

Keeping politics out of AIDS

The third international conference on AIDS last week was as much politics as science. But politicians are proving to be inept at putting their weight where most good would be done.

PRESIDENT Ronald Reagan, who made his first major speech on AIDS (acquired immune deficiency syndrome) on the eve of the third international meeting on AIDS in Washington last week (see page 449), must be wondering why his evident attempt not to offend liberal opinion should nevertheless have earned his representatives boos and jeers from the audience of 6,350. Here is an explanation.

Two years have passed since the first international meeting on AIDS, at which the danger to the US population from an epidemic of the disease was already apparent. The president's reluctance to declare himself sooner can be easily be understood. If all Americans lived in the conservative moral world the president idealizes, that of life-time monogamous heterosexual relationships, there would be little to worry about; a meeting on AIDS might draw the 25 people attending the first meeting on retroviruses 25 years ago.

Yet, whatever his moral views, the president went out of his way last week to stress that there must be no personal discrimination against AIDS victims, ninety per cent of whom are homosexuals or intravenous drug users. Such a clear statement, late though it may have been, is creditable. The president also avoided calling for the draconian programme of nationwide compulsory testing that some of his conservative colleagues are believed to favour. Instead, he urged that compulsory tests should be reserved for aliens seeking residence in the United States and those in federal prisons, while "routine" tests would be offered to (but not forced upon) those seeking marriage licenses and patients at drug rehabilitation centres and venereal disease clinics.

Proposals

In themselves, these proposals are inoffensive. US citizens are unlikely to be outraged that those not yet US citizens should be tested and, indeed, the Senate unanimously voted in this sense last week. (It remains to be seen whether the plan will apply to the estimated four million illegal immigrants from Mexico now being offered amnesty.) So why was the Vice-President, Mr George Bush, booed when he opened the AIDS conference? And why did participants turn their backs on the Secretary of Health and Human Services, Mr Otis Bowen? Because the president's message, for all its inoffensiveness, offered so little to help those who are most intimately engaged in the struggle to prevent the spread of AIDS. Politically, the president made a gesture to the morally outraged who advocate compulsory testing (and who appear incapable of understanding that it is wholly impractical to detect all carriers of the virus) calculated not to stir up too much trouble elsewhere.

No wonder that the AIDS conference seemed itself to be a political occasion. Among the sideshows, the homosexual activists calling themselves the "Lavender Hill Mob" wore white jackets bearing pink crosses in parody of the uniforms (heavy with gold epaulettes) officers of the Public Health Service are being encouraged to wear by the Surgeon General, Everett C. Koop. On the other side were groups demanding the enforced isolation of homosexuals, the resignation of Koop on account of his "condom-mania" and the issuing of an internationally-recog-

nized test-card enabling its holders to find partners in promiscuity. Laughably, the conference organizers found it necessary to distance themselves from the more offensive views recruited by their exhibition salesmen.

Conclusions

If the president's men had listened to the conference, they would have arrived at different conclusions, which are these:

Compulsory testing would be counter-productive. Those most at risk — homosexuals, intravenous drug abusers and their sexual partners, are likely to be driven into hiding by compulsion. Like it or not, only voluntary tests offered in confidence will succeed.

Large-scale testing for low-risk groups, even those about to marry, is a waste of scarce resources while the prevalence of the AIDS virus is so low. Already there is a two-month delay for voluntary AIDS tests in some cities. But there is a case for anonymous surveys to make accurate estimates of the infection rate (to which the administration appears ready to agree). And while the recent British experience points to the difficulty of reaching members of the high-risk groups, is it sensible that the funds Congress has voted for a television campaign of public education should remain unspent?

Intravenous drug users seem to be the chief conduit by which AIDS is spreading into the heterosexual community of the United States, largely because of their habit of sharing syringe needles. How to help these people, already on the margins of society, stop when a shared needle may be token of friendship and when their habit may betoken anxieties even more serious than those of being infected by AIDS? The best hope seems to be to recruit addicts into drug-treatment programmes, a cause to which the president and Mrs Reagan lent their weight just months ago. In the practical politics of the battle against AIDS, the need now is for more sympathetic help for this already unfortunate sub-community.

Homosexuals differ from drug users in having formed self-help groups, whose existence seems to be responsible for the changes of sexual practice recorded at last week's conference. But the trend towards safer homosexual sex is not yet big enough — and the abatement of the rate of reports of new AIDS cases among homosexuals may simply be a sign of saturation of infection by the virus. Education and counselling remain an urgent need.

For conservative moralists, the underlying paradox of the AIDS epidemic is that effective safeguards of the public health requires a sympathetic accommodation with homosexuals and drug addicts. Even the rhetoric of exclusion may make things worse. It would be to ask a lot of President Reagan that he should give full-hearted backing to the social programmes now needed, let alone to sanction public education programmes extolling the benefits of condoms as barriers to infection. Perhaps he should leave matters to his Surgeon General, an evangelical Christian appointed to his post only after a battle with liberal congressmen, who has shown himself in office to be less interested in moralising than in doing whatever is necessary to stop the spread of AIDS. His reward came at a testimonial dinner last month boycotted, to their shame, by his former conservative supporters. □



Tomorrow's tasks

This week's new government must address itself urgently to the needs of its research enterprise.

TODAY'S election in Britain means that there will be a new government in office tomorrow, 12 June. Most probably, it will be a replica of Mrs Margaret Thatcher's administration of the past eight years, but with a smaller majority in the House of Commons. But even if the Labour Party, the largest of the opposition parties, has managed in the day or so before the counting of the ballots to redress the imbalance against it showing consistently in the opinion polls, the opportunities and the problems will be what they have been for the past several years. Britain retains its conviction that there must be some way of turning science into prosperity and is fiscally in better shape than for many years, but still has three million adults unemployed and has demoralized its research enterprise and its institutions of higher education. The pursuit of fiscal rectitude, entirely laudable, is why, since 1980, public support for higher education and (with caveats) for basic research have both declined. The British, government and otherwise, have given little thought to the complementary question of whether the length of the dole queues is related to what has been done to British education.

That is the issue to which the new British government should bend its energies, and more radically than any party has been promising, for there is ample evidence that the dole queues are filled with people who would be at work if their education had been better suited to the world in which Britain aspires to live. The economic history of the past decade makes that plain. Traditional industries, from textiles to steel-making, have been in decline for the four decades since the Second World War for two reasons — traditional manufacturing did not allow a high-wage economy to compete with manufacturers elsewhere, but there has been neither the capital investment nor the technical innovation that would have allowed profitable competition in the niches of the global markets that others have discovered. Two developments in the early 1970s accentuated Britain's economic problems; membership of the European Economic Community made the competition more effective, while the exploitation of North Sea oil, by increasing the value of the British currency relative to others, had the same effect. The result, as the opposition parties complained during the election, is that British manufacturing output has hardly changed since 1979. The real growth of the British economy since then consists largely of North Sea oil and of the service sector of the economy.

This thumbnail sketch is nowhere challenged. Controversy centres on two questions. How, first, to remedy this state of affairs by "creating" jobs? Opposition plans to spend money on the improvement of Britain's infrastructure would take many people off the dole, but more construction can be only a short-term remedy, and could be inflationary as well. So why, more radically, has Britain not experienced the uprush of new manufacturing enterprises, making products that did not exist in the 1950s, which has marked the recent history of countries as different as Italy, South Korea, the United States and — spectacularly — Japan? Shortage of venture capital, once a drag, is no longer a difficulty. There has also been some notable if unobtrusive growth of the kind for which the British yearn — the pharmaceutical and chemical industries have grown on the back of technical innovation, as indeed has the electronics industry. The snag, and one cause of the Thatcher government's discontent, is that growth in manufacture of electronics has not been as spectacular as that of the world market, or of Britain's own consumption of electronic goods.

Why should this be? Eight years ago, the government's diagnosis was that the British economic climate was inimical to entrepreneurs, whence its policy of reducing taxes. Soon afterwards, the government seems to have come to its own interpretation of the legend that Britain is better at invention than

exploitation, whence its twin policies of persuading research councils towards practical problems and of cutting the budgets of the universities to make them more economical and, latterly, to push them into the arms of industry. The logic is seductive, but the consequences are entirely unexpected. The research councils have been so harried by enforced reorganisation and the creeping shortage of funds that their programmes have taken on the appearance of exercises in public welfare. And the universities, once shielded from direct dialogue with the government by the University Grants Committee, now find themselves harried individually from the centre. The University of Aberdeen seems narrowly to be avoiding the prospect of bankruptcy by planning to shed 150 academics. University College Cardiff has been given until today, 11 June, to choose between bankruptcy and a forced merger with a smaller college on terms dictated by the latter. There seems to be little understanding of the effects of these upheavals on the morale of the people who work in the research institutes concerned, let alone of the social and economic function of basic research and higher education.

This is where tomorrow's government must start. The first need is to understand that higher education is a means of producing the skilled people of a kind whose shortage, in Britain, is the chief reason for the lack of successful industrial innovation. Sadly, although the annual output of British graduates is approaching 100,000, there is ample evidence in the economic experience of the past few years that this considerable output does not meet the economic need. The government's explanation, from which the opposition parties do not dissent, is that higher education pays too little attention to industry as such. What about this other possibility, that it is altogether too eccentric that most British degree courses should last for only three years and should ride on a high-school system marked by such specialisation that most school-leavers know either about science, or about the wider world, but not about both? Can it make sense that Britain, bent on the emulation of its successful competitors, should decline to follow them in worrying about the quality of the education it provides for its young people?

Change

To be fair, the previous government had awakened to the need for change by planning for a core curriculum in the schools, while universities and polytechnics have fallen in with proposals for a broader education of school-leavers. Sadly, these changes will be slow to take effect and are, in any case, compromises. Their success has been further compromised by the battle between the high-school teachers and the previous government, exacerbated by the government's improvisation of plans for allowing independence for all schools during its campaign for re-election. Why not remove the canker from the centre of British education by cutting the link between high-school and higher education, letting the former concentrate on educating young people and giving the latter an extra year to educate young adults?

Tomorrow's government must also go back on some of the administrative decisions of the past few years and months. The notion that there will in future be contracts between universities (or even university departments) and a central money-bag called the Universities' (or Polytechnics') Funding Council for the provision of educational services, which is a recipe for making both kinds of institutions less creative, should be quickly abandoned. The new government must also give the research enterprise time to recover from the traumas of the past few years. More money is not required. Instead, there needs to be a respite from the rhetoric of the past few years, based on the assumption that the only worthwhile research is that directed to the improvement of industrial performance. That ministers leap to these conclusions is understandable, which is one reason why the new government should be guided by wiser counsels. Ministers would be more usefully employed in restoring the morale of a scientific community so injured by the past eight years that its capacity ever to recover is now in doubt. □

AIDS drug trials to start amid controversy

- United States follows Sweden in peptide tests
- Others unable to reproduce laboratory data

Washington

WITHIN a few days of a decision by the US Food and Drug Administration to give the go-ahead for clinical trials of a new anti-AIDS drug, peptide T, controversy about the laboratory data on which the decision was based erupted at last week's International Conference on AIDS.

Speaking the day after Dr Candace Pert of the National Institute of Mental Health had added to her group's published data showing that peptide T can inhibit the entry of AIDS viruses into cells, Dr William Haseltine of the Dana-Farber Cancer Center in Boston, claimed that six laboratories, including his own, cannot reproduce the published data. But at least one other laboratory has reproduced the data on the peptide, already being evaluated in Swedish AIDS patients.

The confirmation of Pert's original work, which was published at the end of last year, comes from scientists at Genetic Systems, the Seattle-based subsidiary of Bristol-Meyers; Pert cannot understand why others following her published

protocol should be unable to get the same results. But Haseltine claims exactly that.

A letter from Haseltine and others who have been unable to reproduce any of Pert's data is due to be published by *The Lancet*. Pert is puzzled by the failure of Haseltine and others to discuss their data with her. Had they done so it would have been possible to ascertain whether the protocol had been followed precisely. Most important, says Pert, is that the concentrations of virus and peptide are as in her paper. That is also Genetic Systems' experience. Dr George Todaro says that in his experience the conditions described in the original paper in the *Proceedings of the National Academy of Sciences of the U.S.A.* have to be closely adhered to.

Pert is an outsider to AIDS research but has much experience of working on peptides and their receptors on cell membranes, particularly opioids and their receptors in the brain. Her interest in the infection of brain cells by the AIDS virus led to the synthesis and testing of peptide T, the core amino-acid sequence of which closely resembles a segment of the envelope protein on the outside of AIDS viruses. It is that segment, she believes, that makes contact with the receptor for the virus on cell surfaces, leading to infection of the cell. And it is by competing with the virus for the receptor that peptide T might block the spread of infection.

Concentrating on peptide T, Pert's laboratory has now extended the initial observations to different cells, further refined the peptide structure and devised a more sophisticated test of peptide binding. In the meantime, Dr Lennart Wetterberg of the Karolinska Institute in Stockholm has given the peptide to four AIDS patients and observed some suggestive return towards normality of the nuclear magnetic resonance images of their brains. Rather than yet claim that the peptide is responsible, he will await the outcome of a controlled trial involving 36 patients that has just begun in Sweden.

US trials will be set up by Dr Frederick Goodwin, principal investigator in the submission approved by FDA. It remains to be seen whether the data of Haseltine, which he says are supported by others in US, UK, Swedish, Australian and Dutch laboratories, will give pause for thought. But with the great pressures to find new drugs for AIDS, the FDA seems unlikely to back away from a drug that at worst is likely to be harmless. Peter Newmark

Chernobyl cancellation

London

THE chairman of the USSR State Committee for Atomic Energy, Andronik Petrosyants, has announced that the construction of two more reactors (Nos 5 and 6) at the Chernobyl site has been cancelled. According to the Moscow weekly *Literaturnaya Gazeta*, the decision was taken "by a large majority of votes" at a conference in Kiev, attended by about 60 scientists. In March 1986, just a month before the disaster, the Kiev weekly *Literaturna Ukrayina* had drawn attention to sloppy workmanship and corner-cutting at the No. 5 site. Nevertheless, shortly after the accident, it was announced that construction work would resume shortly and that reactor No. 5 could become operational in autumn 1988. V.R.

Sizewell go-ahead

London

BRITAIN'S first nuclear power station using a pressurized water reactor (PWR) at Sizewell has been given the go-ahead by the Nuclear Installations Inspectorate (NII), three months after government approval.

The Central Electricity Generating Board plans to start the first stage of construction, a deep barrier wall to prevent seawater intrusion costing £8 million, at the end of the month. Safety requirements will also be monitored at the next stages of construction, according to the NII.

Critics of the PWR question the timing of the licence, at the height of Britain's general election campaign. All the opposition parties say said they would abandon the Sizewell B PWR power station K.J.

● A long-term study of the health effects of radiation exposures from the Chernobyl accident will be carried out by Soviet scientists, a meeting of the World Health Organisation and the International Atomic Energy Agency has been told.

Soviet authorities say they plan to monitor over the next 40 or 50 years the health of 135,000 people evacuated from the most heavily contaminated area. "Unlimited funds" have been assigned to the study.

The research is considered internationally important in establishing whether there is a threshold of radiation dosage for the appearance of effects. □

Argentina's nuclear energy

London

ARGENTINA'S National Commission for Atomic Energy (CNEA) has a new chairperson, Mrs Emma Victoria Perez Freyre. Her predecessor, Alberto Constantini, resigned in April, saying that he could no longer continue under the current financial constraints on CNEA. During the interregnum, a 15-day strike of CNEA personnel for increased salary and pension rates brought virtually all projects to a standstill. V.R.

Random survey for AIDS infection

Washington

UNCERTAINTIES over just how many people in the US population are infected with the AIDS virus are to be resolved by a random survey, according to an announcement made by Health and Human Services Secretary, Dr Otis Bowen, after the close of the third international conference on AIDS (acquired immune deficiency syndrome).

Estimates of the rate of spread of the disease and the extent to which it has penetrated the heterosexual population vary widely and reliable data are urgently needed to establish a deadline. Dr Bowen is proposing that 45,000 citizens, selected to sample accurately the entire population, be asked to give blood samples that could be tested for presence of the AIDS virus. Details of how the survey would be carried out have not yet been released but it is certain that participation would be entirely voluntary and anonymity protected. Despite these safeguards the survey is bound to stimulate some controversy. To be effective, the number of people tested has to be large enough to include significant samples of the high-risk groups that include homosexuals and intravenous drug users. Members of such groups may be unwilling to participate. Alun Anderson

Genentech falls at the last fence in race to market heart drug

Washington

THE US Food and Drug Administration (FDA) has strengthened its position as the agency pharmaceutical and biotechnology companies most love to hate with a surprise recommendation from an advisory committee that a heart drug, tissue plasminogen activator, is not yet ready for the market. The crowded hearing at which the announcement was made ended in confusion with representatives of Genentech Inc., the drugs manufacturer, demanding a fair deal, and stock analysts rushing for the door, pulling portable phones from their briefcases as they went. Almost \$1,000 million was wiped off the value of Genentech's stocks and other biotechnology companies were dragged down with it.

Tissue plasminogen activator (TPA) has not been a particularly controversial drug. It is found naturally in the body and activates systems that dissolve blood clots. Genentech's clinical trials show that it does so more effectively than streptokinase, a drug that has been on the market for years and to which the FDA has just given approval for a more direct, intravenous administration procedure. Genentech's presumption was that, following as they were in the trail blazed by streptokinase, evidence of the efficacy of TPA in dissolving clots in the heart would necessarily imply that it would help clear the blocked arteries that cause heart attacks. Clearly, the company believed it had been encouraged in this view by the FDA's Office of Biologics, Research and Reviews.

The FDA's cardiovascular and renal advisory committee thought differently. Perhaps encouraged by a magnificent Italian study for streptokinase, dealt with on the same day, that involved almost 12,000 patients in coronary clinics and showed that the mortality rate was reduced, the committee said that evidence of TPA's clinical effectiveness in prolonging life was needed. The recommendation, although not yet officially accepted by the FDA, sets back Genentech's plans and presents them with a big problem. Clinical evidence that TPA prolongs the life of heart attack patients could only be obtained by trials in which some patients received the drug and others a placebo. But given the evidence now available that TPA clears blocked arteries, most doctors would find the experiment unethical.

The FDA is under some pressure. If time-consuming trials are now begun that eventually prove TPA's effectiveness in helping overcome heart attacks, there is bound to be criticism that the drug was kept off the market for too long. Who will then be to blame for the delay? Ugly (but unsubstantiated) rumours already abound

that part of the reason for the FDA committee's tough stand was that they were under pressure to show who was boss. But at the same time, many industry observers think Genentech was naive to have gone to the hearing without at least a prepared answer to the obvious question of whether dissolving blood clots helps heart-attack victims. The most likely outcome is that the FDA will accept evidence showing that TPA improves left-ventricular function and not demand survival data. But nobody at the FDA or on the advisory committee is prepared to make a statement. That is perhaps not surprising given that the market for TPA could be worth thousands of millions of dollars. Other biotechnology companies, behind Genentech in the race to market TPA, welcome every delay as a chance to catch up. No US patent has been awarded for TPA, and see below for details of its problems elsewhere.

Alun Anderson

Indo-French science cooperation begins

New Delhi

INDIA and France have demonstrated their wish to boost cooperation in science by formally establishing the Indo-French Centre for Promotion of Advanced Research (IFCPAR) in New Delhi, to promote joint research in advanced areas of both fundamental and applied research.

An Indian scientist will direct the centre, which is to be managed by a governing body of five members each from India and France. The cost will be shared, and an eight-member council of an equal number of French and Indian scientists will consider research proposals and training. Although it has some commercial technical collaborations, France has been lagging behind other Western countries in bilateral cooperation with Indian institutions.

At the first meeting of the governing body last week, the centre identified five areas for cooperation: mathematics, informatics, optics and opto-electronics, energy and biotechnology.

K.S. Jayaraman

TPA patent battle rages in United Kingdom

London

TOP guns from both sides of the Atlantic are lining up in London's High Court this week, ready for the opening salvos of Britain's first major biotechnology patent battle.

The British pharmaceutical firm Wellcome is challenging the patent held by US-based Genentech on human tissue plasminogen activator (TPA), a substance that holds great promise for treatment of cardiovascular diseases (see above). The patent, issued in February 1986, is very comprehensive, covering "human TPA, pharmaceutical compositions containing it, processes for making it, and DNA and transformed cell intermediates thereof".

It is this breadth that is opposed by Wellcome, as well as the central issue of novelty. Using recombinant DNA technology to develop human TPA, Wellcome's lawyer told the court, was "a matter of routine" once it was accepted that the substance was needed in quantity. "It is a highly desirable product, so the motive is particularly high," the lawyer said. "You mean the profit motive," the judge replied, adding, "without that we shouldn't all be here today".

Both companies are clearly willing to pump all necessary resources into the court case, which is expected to continue for three weeks, with banks of scientists and mountains of reports cramming the courtroom. It is widely seen as a test case for the patentability of the products of biotechnology.

In the opening arguments, Wellcome's

lawyer pointed out that the patent covers "all forms of TPA except that found naturally in humans, and all ways of making it, known and unknown." He told the court, "TPA is not an invented product — but even if it is, it is quite wrong to claim (in a patent) every other way to get it. Even if you find a new way, there is no way you should have a patent on the old way." He said that the patent would give Genentech rights on the substance for 20 years.

Wellcome argues that standard genetic-engineering approaches would have been able to clone a gene encoding TPA at the time the patent was applied for in 1983, although not necessarily by the same route taken by Genentech.

Genentech is expected to argue that using mammalian cells instead of the usual bacterial cells is an innovation that classifies as an invention when company lawyers begin presenting their case next week. Evidence will also cover the difficulties of handling messenger RNA.

Wellcome told the court that the patent is already being infringed. It is thought that at least 30 biotechnology companies are making TPA, although no one has started marketing it yet. Genentech and its European partner Dr Karl Thomae GmbH, subsidiary of Boehringer Ingelheim, have a major investment in plant, and WelGen, a partnership of Wellcome and Genetics Institute, plan to open a TPA plant by 1989. Scientists are already working on second-generation products, and waiting for the outcome of the case.

Kathy Johnston

Taste of power brings conflict to West Germany's Greens

Munich

WEST Germany's effervescent Green Party of the early 1980s has grown into a troubled adolescent. Accepting responsibility, the Greens are discovering, is not nearly as easy as challenging those in authority. Some of the party's naive idealism has been lost, replaced by bitterness.

This has been a rollercoaster year in German — and Green — politics. In January, more than three million voters gave the Greens 8.3 per cent in the *Bundestag*, their highest total ever. In the 17 May election in Rhineland-Palatinate, the Greens swept into an eighth *Landtag* (state parliament), an all-time high.

The election in Hesse on 5 April, when the Greens polled a record 9.4 per cent, should have been another reason to celebrate; they were even thinking of demanding a second cabinet portfolio in West Germany's only Social Democratic (SPD)/Green coalition. But the SPD collapsed and the conservative Christian Democrats (CDU), led by Federal Environment Minister Walter Wallmann, formed Hesse's new government in a traditionally SPD bastion.

This was a nasty slap in the face for the Greens, who had been campaigning for the closing of two nuclear-fuel processing companies in the city of Hanau. The government of Hesse had even challenged in the courts the federal government's jurisdiction over the plutonium industry (which suit Wallmann has now withdrawn).

The election result in Hesse has had wider consequences. In Hamburg on 17 May, it seems that thousands of voters deserted the Greens for the SPD to head off a further victory by the CDU. At the same time, sniping between two rival Green factions that had begun after the Hesse election escalated into a war on 3 May, when three so-called "Fundis" were elected at a national party gathering to head the unpaid 11-member *Vorstand* (executive committee) of the Green party. This election left the "Realist" faction of the party, known as "Realos", in distress.

The internecine debate has produced rhetoric as shrill as that heard from the Greens after Chernobyl, but they are now criticizing each other. "Realo" Otto Schily, a member of the *Bundestag*, has accused "Fundi" Jutta Dittfurth, the new Executive Committee speaker, of having an "insanely distorted image of reality" because she allegedly likened German census-takers to fascists.

Dittfurth says she simply meant that there is no such thing as harmless data. Censuses taken under the Nazis revealed the identity of Jews, and the 1987 census

forms also ask a question about religious affiliation. Dittfurth accused Schily and his allies of wanting to split the party.

The internal battle centres on differences of opinion about accepting power. The Fundis reflect the roots of the Green Party, which began in the late 1970s as a grass-roots environmental movement, and advocate refusal to participate in an unjust and "inhuman" society. They demand a "restructuring" of society that will eliminate pollution, nuclear weapons and large corporations.

The Realos accept these ideals, but advocate very different means of achieving them. Their immediate goal is



Eight Länder (wheat symbol) now have parliamentary representatives from the Greens.

to prevent right-wing governments from taking power and score visible political gains wherever possible. Barbara Rust, another *Bundestag* Green, says for example that shutting down some West German nuclear power plants would be acceptable if it were impossible to shut them all. She has criticized the Greens' conduct of the Hamburg election for not offering plausible conditions for a coalition with the SPD, saying that "Our supporters expect us to use every option available to us".

The Fundis disagree. Dittfurth reiterated after the Hesse election her earlier view that success at the polls is "not terribly important". According to Christian Schmidt, also on the Executive Committee, the Greens do not want to grow up.

The Greens, once proud to call themselves "the anti-party party", have become much like the others. Green *Bundestag* members are still paid less than others; the difference is put into an Eco-fund for environmental projects. But the

Greens no longer rotate their *Bundestag* seats as was originally intended.

Political experience has changed much; some Green politicians (a phrase that would itself have seemed self-contradictory four or five years ago) seem to enjoy their positions of power. Yet they lack policies that are not simply criticisms of the other parties. Their anti-nuclear philosophy suffered a telling blow when, after Chernobyl, virtually nothing in the West German nuclear industry changed. "We knew that 80 per cent of the population wanted to switch off nuclear power", says Rust. "We have to ask ourselves if we've done something wrong."

The feud has created an "unaffiliated" wing of the party, called by one newspaper the "Neutralos", including roughly a third of the 44 Green *Bundestag* members. Led by Rust, the new group wants to distance itself from the battle.

But it offers little input as to how to broaden Green horizons. At worst, it seems just an "alternative within an alternative," not knowing just what it stands for. The danger for the Greens is that, having thrived on controversy, they may now turn to conflict within their own ranks on which to feed.

In what direction must the Greens go to maintain what little parliamentary influence they have? (Cold-shouldered by the other parties, they have sponsored only one successful bill — to ban turtle-meat imports — in the past four years.) A survey by the polling company INFAS showed that over 90 per cent of Green supporters in Hesse and Hamburg would have supported a coalition with the SPD, to which the Fundis are opposed.

Another INFAS survey has shown a steady downward trend of support among voters between 18 and 25 which Rust attributes to the Greens' refusal to come to terms with modern technology and so to offer young voters the prospect of finding jobs. Rust would seek to "convert current technologies" into less ecologically harmful jobs to bridge these demands.

But Green *Bundestag* member Christa Nickels is convinced that the recent spate of character assassination "scares people off," as do the radical slogans of the Fundis. Rust also faults the Realos for not taking intraparty disputes seriously enough, perhaps because they are too concerned with winning majorities. Now the best the Greens can hope for, it seems, is that the feuding will be stilled so that they can use the summer to prepare for the September elections in Schleswig-Holstein and Bremen.

Fundi Schmidt criticizes the Greens in general when he says that "what's missing here is the courage to practise alternative politics". But compromising also requires courage. The future for the Greens may depend on outgrowing its belligerent adolescence.

Steven Dickman

Supertrees from Australia set to dominate the world?

Sydney

AUSTRALIAN "supertrees" are about to embark on a wave of world settlement. Having evolved to cope with the continent's arid environment, Australia's flora is among the toughest in the world, and the supertrees about to go overseas are the toughest of the tough. Trees are being developed with extraordinary capabilities such as salt tolerance, drought resistance, resistance to waterlogging, high rates of water transpiration and high growth rates.

A trial batch of salt-tolerant Red River Gums is already growing in Kuwait. Middle Eastern countries need extensive landscaping programmes around population centres, but there is hardly any rain and the little fresh water available is used for drinking. The only water available for plants is saline, so salt- and drought-tolerant trees are in demand.

The supertrees have been developed by collecting seeds from the healthiest trees growing in the harshest conditions throughout Australia, such as around salt lakes and along estuaries. Germinated and grown in glasshouses, the trees are subjected to increasingly severe conditions until only a few survive. Using tissue-culture techniques, these survivors are then cloned for further trials and commercial mass propagation. Culturing and cloning ensures that each sample is genetically identical and so carries the desired characteristics of the parent plant. Selling the

cloned saplings also eliminates the long wait for the trees to produce seeds.

Scientists from several universities and research laboratories are involved in the programme, including the mining company Alcoa, whose interest arises from its commitment to the reforestation of mine sites. According to Dr Victor Harvey of the Commonwealth Scientific and Industrial Research Organisation's Division of Forest Research, tolerance to water with salt concentrations up to twice that of sea water have been achieved.

Scientists are also working on further improvements to the performance of the trees by inoculating them with specific plant-root fungi (ectomycorrhizas) that live in a symbiotic relationship with the trees and can enhance their ability to resist environmental stresses such as salinity and improve their ability to take up nutrients from the soil, leading to faster growth.

This latest wave of Australian migrants will include various species such as the wattle (*Acacia*) and the Australian sheoak (*Allocasuarina*) as well as the common eucalypt. This programme should answer criticisms often levelled at the eucalypt by farmers in developing countries, that their foliage cannot be used as animal fodder. The fast-growing trees will be valuable in many of the developing countries as sources of firewood, for re-forestation and for production of pulpwood.

Charles Morgan

Austro/Czech alarm over nuclear safety

London

AN apparent error in translation has caused considerable anxiety about nuclear safety on both sides of the Austrian/Czechoslovak frontier. Mention in the Slovak daily *Pravda* of "discrepancies" between the Soviet designs for nuclear power station and the actual construction of such stations in Czechoslovakia was rendered into German as "defects", "errors" or "breakdowns". The *Pravda* article described the visit of a group of Soviet experts to the Czechoslovak nuclear-power stations where further units are under construction, and said that in checking the construction against the blueprints, 135 discrepancies were noted at Jaslovské Bohunice, 221 at Dukovany and "several more" at the Mochovce construction site.

The *Pravda* article appeared on Friday, 8 May. By the following Monday the Austrian media had given wide publicity to the supposed "defects". Two days later, the Chief Inspector of the Czechoslovak nuclear safety authority Jiri Beranek convened a press conference at the Czechoslovak embassy in Vienna to explain what was meant by "discrepancies". Nuclear power stations in Czechoslovakia, he said, are built to Soviet designs but to Czechoslovak technical and safety standards, which in many cases are stricter than the Soviet ones. Standards for electrical installations, in particular switches, circuit-breakers and colour-coded cables, differ considerably. Other discrepancies refer merely to a slight difference in siting a door or window. Nevertheless, because a Soviet blueprint is used, it is necessary to draw up a special protocol for each deviation and to apply appropriate tests and acceptance procedures. These changes are naturally reported to the Soviet design team, although the latter act only as advisers and no Soviet experts act as inspectors at any stage of the construction or operation of Czechoslovak nuclear power stations.

Unfortunately, by this time, Austrian television had carried the story of the alleged "defects", and this had been seen in western Czechoslovakia, causing further alarm. Beranek had to repeat his explanations for the domestic audience, and the Czechoslovak Minister of Fuel and Power, Vlastimil Ehrenberger, made a similar statement in the federal assembly. Czechoslovak fears, however, have not been fully allayed, and Beranek has since found it necessary to give yet another press conference and radio interview on nuclear safety.

Vera Rich

NASA welcomes whistle-blowers

Washington

ALMOST eighteen months after the Challenger disaster, the National Aeronautics and Space Administration (NASA) is still trying to come to grips with the acknowledged failures of its safety review procedures. Officials announced last week that an independent body will be set up to screen expressions of anxiety, concerns and complaints before passing them on, anonymously, to NASA's own analysts.

Under the new system, which is modelled on a scheme developed by the Federal Aviation Administration (FAA), employees of NASA and its numerous contractors will be encouraged to communicate their worries to an independent facility at the Battelle Memorial Institute in Columbus, Ohio. Cranks and hoaxers will, it is hoped, be sifted out, and genuine concerns will go on to NASA for action.

The advantage of this system, according to NASA, is that it provides a regulated mechanism by which safety problems can be brought up in confidence and formally dealt with. Hitherto, anonymous com-

plaints have not always been treated seriously, while the fear of retribution dissuaded many from attaching their names to letters. Problems stayed hidden, a state of affairs which was largely responsible for the Challenger explosion.

The new procedure will initially cover only space shuttle and orbiter operations, but should spread to all of NASA's departments. Although the shuttle accident led, eventually, to a recognition of organizational problems within NASA, it did not by itself produce solutions. Many employees reacted by "going into a cocoon", as one NASA spokesman put it. The changes now being put into effect are supposed to encourage a more open atmosphere within the agency so that problems may be handled openly and efficiently.

For the time being, at least, the Battelle Institute has no responsibility beyond notifying NASA of a problem. Anyone who wishes to ensure that a complaint has not become lost in the bureaucracy can check on progress only by abandoning anonymity.

David Lindley

US attitudes to biotechnology show qualified support

Washington

ENTHUSIASM, concern and ambivalence are all part of the US public's attitude to biotechnology, according to a new report by the US Office of Technology Assessment (OTA)*. Although 66 per cent of Americans believe that genetic engineering will improve life for everyone, 52 per cent think that improvement may be accompanied by dangers to people or the environment.

The report's findings are the result of a study by the Louis Harris and Associates agency, which sampled the views of 1,273 citizens last autumn. Most of the US population thinks it understands the concept of genetic engineering. Although 62 per cent of the public feel that the benefits of new technological and scientific developments outweigh the possible risks,

the OTA study reveals a relatively widespread belief that biotechnology is inherently dangerous. But fears about biotechnology are generally vague and ill-defined, prompting OTA to conclude that the public's beliefs "are not necessarily based on factual information". On the other side, about half of those responding to the survey feel that the risks of genetic engineering have been exaggerated.

Most people are willing to take on relatively high rates of risk to reap the benefits of genetic engineering, the study finds. Slightly over half would approve the use of a genetically engineered product that would significantly increase farm production if the risk of losing a local species of plant or fish were 1 in 1,000. But the amount of risk must be quantified for most people to approve the use of a biotechnological product in the environment.

Resistance to field trials of new organisms was not as high as critics of biotechnology have suggested. A surprising 67 per cent said they would either approve or not care if a genetically engineered product were field-tested in their community. But people would be more hesitant to approve the dissemination of a biotechnological product on a large scale.

A consistent pattern revealed by the report is that the US public supports potential gains from genetic engineering, but is uneasy with the abstract idea of genetic manipulation. Nearly half the public feel that it is morally wrong to alter the genetic make-up of human cells, but about three-quarters would approve of specific therapeutic applications to reduce someone's risk of developing a fatal disease later in life or to cure a usually fatal disease. Most would even favour genetic therapy on germ-line cells to prevent potentially fatal genetic defects from being passed on, even though the scientific community has been reluctant to open this Pandora's box.

The study uncovers a problem for the government: on the one hand, the public believes that the government is best suited to make regulatory decisions to protect public safety; but on the other, it feels that university scientists, public-health officials or environmental groups are better sources of information on the subject. When asked if they would believe a government agency's assurances of safety over the warnings of an environmental group, most chose the environmentalists. This lack of faith in government agencies may reduce the government's role in public-education efforts. **Carol Ezzell**

*New Developments in Biotechnology: Public Perceptions of Biotechnology. OTA-BP-BA-45. US Government Printing Office, Washington, DC.

New education satellite planned for Europe

London

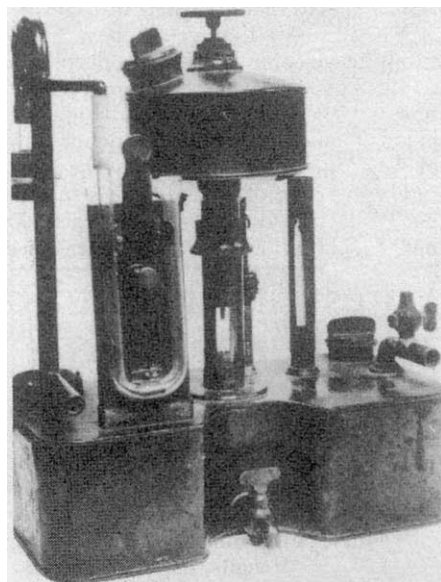
A SATELLITE network linking the centres of excellence in European science teaching is to be launched at the beginning of next year for use in industry and academic institutions. The project, backed by the European Community and five of the principal electronics groups in Europe and the United States, will provide a unique vehicle for advanced education through 'remote learning'.

The network, called PACE, the Programme of Advanced Continuing Education, will offer students at universities or in industry courses that can be credited toward an advanced degree or post-doctoral qualification. It will be modelled on the American National Technological University (NTU), founded in 1985. The programme, which will chiefly benefit students or employees in the applied sciences, has received financial and technical support from companies throughout Europe and the United States. Typically, the network would offer courses on software engineering, microelectronics, materials sciences, telecommunications and artificial intelligence. The service is expected to be carried on a European telecommunications spacecraft and to be given full support by the respective national telecommunication authorities. According to the PACE organizers, the chief suppliers of services or courses to the network will be those in European universities and other higher-educational and scientific institutions "recognized of top quality and as leaders in the field concerned, and the users will be engineers, managers, technicians and scientists who need advanced training and specialist knowledge. The organizers have already established cooperative links with European bodies representing higher education and the professions.

The group will also try to establish agreements in the United States. Although the planning is at an early stage, collaboration with NTU is typical of the links being considered, allowing American postgraduate expertise to be offered to European students.

Britain's Open University, which offers courses to 110,000 students through the public broadcasting network, is finalizing agreements for six programmes, candidates for broadcast across Europe, to be used as pilots to sell its services. The experimental satellite Olympus, scheduled for launch next year, has also attracted the attention of the Open University. An educational channel is to be offered free to suppliers by the European Space Agency. **Bill Johnstone**

Anaesthetics display



SHIPWAY'S warm ether/chloroform intratracheal apparatus, on display at the exhibition "No Laughing Matter: Historical Aspects of Anaesthesia" at the Wellcome Institute for the History of Medicine, 183 Euston Road, London, UK, which opened on Monday this week. The exhibition uses rare archival material to show how the discovery and development of anaesthetics caught the interest of the public and demonstrates the contribution of anaesthetics to medicine and research. In Shipway's apparatus (1916-35), a copper water bath held a waisted chloroform bottle and a vaporizing chamber for ether. Air, with or without oxygen, was passed through the anaesthetic agents by a motorized pump. (Actual size 255 × 286 × 285 mm.) □

Chirac's privatization programme reaches French science funding

Paris

LATER this month a new bill, which aims to stimulate private patronage in a wide range of areas, from science and the natural environment to culture and sport, goes before the Senate (the French parliament's upper house). The "Patronage Bill" (*Projet de Loi de la Mécénat*), put forward by Prime Minister Jacques Chirac and Minister of Economy, Finance and Privatization M Edouard Belladur, offers the now familiar carrot of increased tax incentives to private enterprises willing to donate part of their profits to recognized charitable trusts.

Sensitive to accusations that the bill is merely a device to camouflage cuts in spending on science and the arts, the government, in a senate report, insists that by seeking to stimulate private

patronage, it is not shirking its responsibilities. An increase in the tax threshold certainly looks good on paper. If full use were made of the law, the cost to the government in lost revenue from taxation is estimated at 900 million francs (£90 million). But figures provided in the report suggest that tax incentives may not be a very effective way to raise money. The new law will allow 0.3 per cent of gross turnover to be deductible from tax. But under the existing Finance Bill, private enterprises can already deduct up to 0.2 per cent of gross turnover by giving to charitable trusts — and full use is not made of this concession. At national level, donations in 1986 represented only about 0.0001 per cent of total turnover.

Although the main thrust of the patronage bill is aimed at the arts, medical

charities will also benefit. One of these, the *Fondation pour la Recherche Médicale*, was spurred by the prospective bill to carry out a large-scale fund-raising drive and has already received positive replies from a number of major companies. Air Liquide, France's major manufacturer of industrial gases and parent company of a dozen subsidiaries, announced this week that it is to give FF6 million (£600,000) to the foundation over the next three years. A memorandum issued by the company said that the money will be used "to stimulate progress in our understanding of the absorption of gases by the body . . . their transport, normal and pathological action and their use in biology and medicine".

Exactly how the money will be spent, within the terms of the donation, will be decided by the scientific advisory board of the foundation. M Gerard de Chavnac Lanzac, president of the foundation, said it was likely that up to 60 per cent will provide exchange bursaries for young researchers. Last year the foundation awarded 144 grants for French researchers to travel abroad and 50 to enable foreign researchers to visit French laboratories.

Although Air Liquide's donation will surely be welcome, meriting a full column of the right-wing daily *Le Figaro*, it cannot be considered over-generous within the provision of either the existing finance law or the proposed patronage bill. The gross turnover of one of the company's subsidiaries, Lippa, alone was over FF1,700 million (£170 million).

Peter Coles

India plans top-level spending on superconductor research

New Delhi

INDIA's prime minister, Rajiv Gandhi, has set up a Cabinet-level committee under his own chairmanship to promote research related to ceramic superconductivity. This development is a measure of India's confidence that its researchers have much to contribute to (and that India stands to gain something from) the international excitement over the new phenomenon.

The members of the new panel include the ministers of science, finance and human resource development, the cabinet and finance secretaries and the heads of the science agencies supporting research on superconductivity. The committee's members also include Professor M. G. K. Menon, Gandhi's science adviser, Professor Yash Pal, chairman of the University Grants Commission, and four prominent industrialists.

Dr Vasant Gowariker, secretary of the Department of Science and Technology (DST), a member of the panel, says that the formation of the panel under the prime minister is a sign of the government's commitment to research in superconductivity. Gandhi has also set up a Programme Management Body under Professor C. N. I. Rao, the director of the Indian Institute of Science and Technology at Bangalore and chairman of the government's science advisory committee, to coordinate research at government and industrial laboratories. This body will have the executive power and financial muscle to pursue the project to its ultimate goal. Gowariker says that the equivalent of several million dollars has been allocated

to the coordinated programme.

Seven Indian research groups are engaged in superconductivity research: the Indian Institute of Science (Bangalore), the Tata Institute of Fundamental Research (Bombay), the Bhabha Atomic Research Centre (Trombay), the Madras Indian Institute of Technology, the Indira Gandhi Centre for Atomic Research (also at Madras), the National Physical Laboratory (New Delhi) and the National Chemical Laboratory (Pune). Each of these laboratories claims to have produced ceramic material superconducting up to 120 K, and the National Physical Laboratory claimed room-temperature superconductivity just a week ago.

Despite excitement about the potential of the field, Gandhi's panel will have to talk tough to make the groups work together. The DST, referring to the "varying levels of accomplishment" at the different laboratories, says that these efforts "will come to nothing" unless there is a purposeful direction and a pooling of talent.

Unhappily, the atmosphere is already a little sour. Vulgar quarrels have erupted among the research groups, which are rivals. For example, the Bombay and Trombay groups, originally partners, have already split up. One scientist at the National Physical Laboratory asks how, if we have "rivalries within this institute", we can "expect cooperation on a national scale".

K.S.Jayaraman

● The formula for the room-temperature superconductor was cited wrongly last week: the correct stoichiometry is $Y_2(Ba,Sr)Cu_3O_x$.

Forest fire threatens Manchurian tiger

London

CHINA's voracious appetite for timber is unlikely to be dulled by the huge forest fire which blazed over 3.7 million hectares of Manchuria. Scientists fear that the loss of timber trees may mean enormous pressure on Chinese authorities to allow logging within internationally important nature reserves containing many rare species, including the Manchurian tiger.

The fire at one point threatened the Huzhong Nature Reserve, the third largest reserve in China, but the mobilization of more than 50,000 firefighters and a programme of cloud-seeding finally brought the blaze under control.

The stands of larch north and west of the Huzhong Reserve that were wiped out are relatively unimportant as far as rare species are concerned, but the timber available for China's rapid industrial expansion has been greatly reduced.

Scientists say that if logging is allowed in nature reserves, the Manchurian tiger could be exterminated; each of the 100 or so animals still in the wild needs more than 20 km² of forest to survive. Kathy Johnston

Soviet writers protest successfully about environment science

London

THE campaign to avert the southward diversion of the Siberian rivers has not yet been won, a leading Soviet author, Sergei Zalygin, has warned his fellow writers. Addressing a plenary board meeting of the Union of Writers of the USSR, he warned that the decree halting work on the diversion had authorized "further study" of the problems involved. "And so, if in the past a design organization received money for the project, now it will aim to get it for both the project and the problem", he said.

Zalygin is a member of the group of Russian writers known as the "villagers", which is widely credited with a decisive role in halting the diversion. Although the role of the villagers has been tactically denied by Academician Aleksandr Yanshin, a leading geographer, who has publicly stated that the cancellation was based solely on "expert opinion", Zalygin and his colleagues have come under attack from the hydrotechnical establishment.

It is not merely a matter of redeploying those who worked on the project — 6,000 specialists in just one "institute for diversion" — Zalygin said. Some of these engineers and designers had lived with the project all their working lives and had received medals, prizes and titles of honour for their efforts. The psychological effect of being told that all this work was "to no purpose" was considerable.

The writers expressed concern about the drying up of the Aral Sea, the Daugavpils hydroelectric scheme in Latvia, and the Kara-Bogaz-Gol dam — all cases where central planning has paid

insufficient attention to local consequences. The Aral Sea has been depleted by over-use of its feeder rivers for irrigation (a situation that the diversion project was intended to rectify), the damming off of the Kara-Bogaz-Gol bay from the Caspian to reduce evaporation losses of the latter destroyed the brine feedstocks of the chemical industry of Turkmenia,



and the Daugavpils hydroelectric scheme is viewed as a threat to the environment and population of Latvia and Byelorussia. All these controversies could be interpreted as a protest by the non-Russian republics against Moscow; hence the speakers at the writers' meeting were careful to stress that in taking up the ecological needs of their republics, they relied on the example and support of their Russian colleagues.

In spite of what Zalygin sees as the

Anger over nuclear waste plans

London

OFFICIAL British government advisers on nuclear waste disposal are angry because they were not consulted over the government's recent decision to abandon four potential trench disposal sites, and have demanded a meeting with the Secretary of State for the Environment to discuss the issue. The acting chairman of the Radioactive Waste Management Advisory Committee (RAWMAC), Professor John Greening, has warned the British public not to be misled into thinking there is independent assessment of the government's nuclear-waste disposal policies. The committee, appointed by the Secretary of State for the Environment to offer independent advice, first heard of the decision through the media and has still not been given enough information for the proper assessment of the change in nuclear waste disposal plans.

"There is a fair degree of annoyance" among committee members over the way the decision was made, Greening says. "We're puzzled that after all the research we had done we were not consulted." He suggests there was "some political advantage" in the decision.

RAWMAC is still concerned about whether the Drigg repository can handle all the low-level waste produced in Britain before an alternative deep repository site is developed. "Some calculations of committee members don't coincide with those from NIREX [the Nuclear Industry Radioactive Waste Executive]. The Drigg site may fill up faster than they think", Greening says.

Kathy Johnston

United Kingdom to process in parallel

London

BRITISH public agencies have announced a £3.5-million programme to encourage the use of the microchips called transputers in novel products. The programme is being supported by the Science and Engineering Research Council (SERC) and by the Department of Trade and Industry.

Transputers are the most recent microchips produced by the design and manufacturing company Inmos, founded with financial help from the British government but now a subsidiary of Thorn-EMI, the British consumer products manufacturer and defence contractor. The devices, which concentrate memory, processing capacity and communications channels in a single silicon chip, are reckoned to be especially promising as components in the design of parallel processing computers.

Inmos is supported by others in its belief

that its transputers are at present more advanced than comparable devices manufactured elsewhere, but there is anxiety that the present advantages are being eroded by developments in the United States and Japan.

This point was put forcefully to SERC, at the end of last year, by its Engineering Board's computer facilities committee, which argued that the market for hardware for parallel processing could amount to several hundreds of millions of pounds a year, but that effort was needed to ensure a fair share for transputers.

The new programme will be coordinated from SERC's Rutherford-Appleton Laboratory and will seek both to promote the use of the device in products now being developed and to tap academic expertise in the development of software for use with transputers.

Bill Johnstone

vested interests of the engineers, *glasnost* is producing high-level measures on environmental issues. During the past few weeks, government and party resolutions have been passed on the protection of Lake Baikal, and a number of high officials have been censured, dismissed or allowed to take early retirement for slackness in implementing previous legislation. Similar measures have been taken to step up protection of Lake Ladoga (which, like Lake Baikal, is under considerable threat from the cellulose and the paper industries).

Last week a "public committee" for the Aral Sea and a "Soviet Greenpeace" were established under Zalygin's chairmanship, as a commission of the Soviet Peace Committee "to combine the anti-war struggle with environmental protection". And the Minister of Forestry of the Turkmen SSR has been dismissed for "inertia" in protecting the nature reserves and national parks under his jurisdiction — the highest official, so far, to be ousted on environmental grounds.

Vera Rich

Science versus engineering

SIR—Whereas you ask (*Nature* 326, 737; 1987) why technology is so isolated from science, I prefer to wonder why science is so isolated from technology and what role both may have in the engineering to which you address your views. I have previously covered the issue in an article that fully¹ agrees with your differences in motivation and reward between scientists and technologists. One recalls Alvin Toffler's definition that technology was developed to meet one of two criteria: "Does it make a buck or a big bang?" For it is the commercial aspects that divorce the scientists from technology and its applications in engineering.

Most (electrical) engineers are not interested as such in descriptions of new semiconductor devices, laconic or not. It is the ability to get into economic and reliable production those devices that satisfy an identified need, and to incorporate these devices into products the customers want, that drives the engineer. Most new products use only existing technology and it is in that light one must place the work of engineers, technologists and scientists. A spectrum does exist from science and applied mathematics to engineering, but the research role as understood by the scientist (and too often the politician) is a small one in the successful operation of a strong, internationally competitive manufacturing or consulting engineering enterprise.

In Britain, as in Australia, the scientist is too divorced from commercial aspects of technology and at the same time does not appreciate the professional nature of engineering in which "design for profitable manufacture should be as much the focal point for engineers as the human body is to the medical profession"². If external validation of the quality of graduate engineers is considered a mystique, so be it, but, as in medicine, the engineer has responsibilities for, and suffers the consequences of, engineering disasters in a way that the scientist does not. Understanding the difference between an academic science and a profession like engineering is a problem not yet resolved within the universities, our society and the scientific literature.

In engineering publications one sees the damage a scientific attitude to information and knowledge can do to industrial competitiveness. Technology, whether 'scientific' or empirical, in materials or processes or devices, is perhaps the major contribution to commercial success of the manufacturer, consulting engineer or whatever. It is a fatal fault to forget this by too generous publication or gift. It is said that Japan paid only \$15,000 million towards the intellectual property it

needed to achieve its current pre-eminence in engineering manufacture. The rest it obtained from a too-generous approach to detailed publication and only more recently from its own developments. To obtain details on advances in technology (perhaps in other than laconic style), I recommend ignoring *Applied Physics Letters* and looking at the 400,000 allegedly new inventions being protected by patent application each year.

That the editor of a scientific journal considers one to be venal, able to be bribed, to be sacrificing principles from sordid motives, if one patents and exploits intellectual property or technology, provides the best example yet that technology, let alone engineering, does not need science.

T.W. COLE

(on leave from the University of Sydney and the Warren Centre for Advanced Engineering)

St John's College,
Cambridge CB2 1TP, UK

1. Cole, T.W. *Proc. R. Society New South Wales*, **118**, 121-6 (1985).
2. Crossland, B. *Proc. Instn Mech. Engrs* **201**(B1), 1-10 (1987).

Citation counts

SIR—Recent letters in *Nature*^{1,2} on the use of citation counts as evaluative measures for allocating resources prompt us to comment on some of the data problems of citation analysis.

Citation analysis has been in existence for more than a quarter of a century; its literature now totals about 3,000 publications^{3,4}. Yet its basic assumption — that scientists cite (reference) what influences them — has not been tested^{5,6}. No-one has read a single paper and compared the information in it with items in its bibliography.

In a series of studies, we read papers in fields with which we are familiar and compared the information in the text with what was referenced. We found that only 30 per cent of the influence was cited⁷. We also found that citing is highly biased⁸. Some influences were always cited when used and others were never cited. About a third of the citations were to secondary sources, which means that a third of the credit that was given went to someone other than the originator of the idea or discovery⁹. Instead, they either do not cite at all or give positive and negative credit simultaneously. Brooks examined citer-motivation and found that 71 per cent of references were multiply motivated, a finding that undermines simple motivational models of citing^{10,11}.

We have recently completed a review of the data problems of citation analysis and find that only about half of them have been studied at all and that none of them has been studied exhaustively¹².

Add to these problems those discussed

by others^{2,6,10,13-17}, and it becomes clear that it is premature to use citations for evaluative purposes. Before such a step can even be contemplated, we must know a great deal about citing behaviour, which requires that we study not only data problems and citer motivation but also the personal interactions of scientists by which information is exchanged and 'negotiated'^{13,18}.

M.H. MACROBERTS
B.R. MACROBERTS

740 Columbia,
Shreveport, Louisiana 71104, USA

1. Hall, G.G. *Nature* **325**, 478 (1987).
2. Johns, G. *Nature* **326**, 238 (1987).
3. Smith, L.C. *Lib. Trends* **30**, 83-106 (1981).
4. Vlady, J. *Czech. J. Phys.* **B35**, 1389-1436 (1985).
5. Cole, J.R. & Cole, S. *Science* **178**, 368-375 (1972).
6. Cozzens, S.E. *Int. Soc. Sociol. Knowledge Newsl.* **7**, 16-21 (1972).
7. MacRoberts, M. & MacRoberts, B. *Soc. Stud. Sci.* **16**, 151-172 (1986).
8. MacRoberts, M. & MacRoberts, B. *J. Am. Soc. Inf. Sci.* (in the press).
9. MacRoberts, M. & MacRoberts, B. *Soc. Stud. Sci.* **14**, 91-94 (1984).
10. Brooks, T.A. *J. Am. Soc. Inf. Sci.* **36**, 223-229 (1985).
11. Brooks, T.A. *J. Am. Soc. Inf. Sci.* **37**, 34-36 (1986).
12. MacRoberts, M. & MacRoberts, B. *J. Amer. Soc. Inf. Sci.* (in the press).
13. Edge, D. *Hist. Sci.* **17**, 102-134 (1979).
14. Gilbert, G.N. *Soc. Stud. Sci.* **7**, 113-122 (1977).
15. Bavelas, J.B. *Can. Psychol. Rev.* **19**, 158-163 (1978).
16. Chubin, D. *Scientometrics* **2**, 91-92 (1980).
17. Swales, J. *Appl. Linguistics* **7**, 39-56 (1986).
18. Collins, H.M. *Soc. Stud. Sci.* **4**, 165-185 (1974).

Turin shroud

SIR—I first wish to assure Denis Dutton (*Nature* 327, 10; 1987) that all the institutions involved in the proposed radiocarbon dating of the Shroud of Turin are fully aware of the crucial need to ensure that the 'chain of evidence' remains unbroken. It was to meet this need that the British Museum accepted the invitation to act as 'guarantor' and independent observer.

The purpose of the meeting in Turin last autumn was to devise procedures for every step of the sampling and testing, procedures which could and would be monitored at every stage by the three certifying institutions, the British Museum, the Pontifical Academy of Sciences and the Archbishopric of Turin, to preclude any possibility of tampering with the samples.

These procedural steps have yet to be finally agreed by the Pontifical Academy of Sciences and the Archbishopric of Turin so I am not at liberty to divulge their details. But, I can reassure Dutton that should the proposed procedures be amended to introduce a possibility of tampering with the samples, the British Museum would decline to act as a certifying institution. Nor would the radiocarbon dating laboratories then necessarily be willing to participate in the project.

M.S. TITE

Research Laboratory,
British Museum,
London WC1B 3DG, UK

A new development technology?

The Rockefeller Foundation of New York is seeking new ways of adapting science to the needs of the world's poorest nations. Its problem is that technology is necessary but not sufficient for improvement.

WHAT can science, in the hands of a private foundation, do for the huge cause of the more effective economic development of the poor countries of the world? This is the conundrum that the Rockefeller Foundation has set for itself and which, at the end of last month, it posed to a mixed group of people as if it were a question in an elegant examination paper. As with the best examination papers, there were comprehension tests as well — first-hand accounts of past successes, public health in Costa Rica, agriculture in the Punjab of India and the like — from which people were challenged to draw general conclusions.

Why the Rockefeller Foundation? The simple answer is that the foundation stands out from the general throng of private foundations by its record in this field. Dr Normal Borlaug's wheat-breeding station in Mexico, for example, was originally supported by the foundation until it became the hub of the network of international agricultural research institutes supported chiefly (but not entirely) by the World Bank. Much earlier, there had been the successful war against yellow fever, waged with Rockefeller money. Now the foundation is openly looking for similar tasks on which to spend between \$50 and \$60 million a year.

The trivial answer is even simpler. One of the foundation's more enviable assets is its delectable villa at Bellagio, the village at the tip of the peninsula separating the descenders of λ -shaped Lake Como in Northern Italy, which has been closed for the past year so that it could be made even more comfortable. Old hands now say that the new villa is distinguishable from its former self only by the absence of cracks in the plaster and the reliability of the plumbing. There is rejoicing at the reopening of the villa among those scholars who believe they may one day be able to make a case for a spell in residence, especially because there is now room for a couple of dozen of them at a time.

To the obvious jibe that such a place is too far distant in character as well as geography from the poor countries of the world to be a suitable location for a discussion of foreign aid, there is this riposte: Bellagio, strategically placed across one of the routes into Northern Europe, is surrounded by such visible evidence of Italian history, ultimately rooted in Europe's once-poor societies, that people cannot but be moved by the demonstration of the

feasibility of development, slow though it may be.

The foundation's own answer to the question of why it should be engaged, the basis of its board's decision a year ago to look for new initiatives in the application of science to development, is partly that it has the experience (and the network of field officers in developing countries) and partly that the use of private funds in these connections can be "emboldening". But, properly, the foundation acknowledges that times have changed, and that the resources it can deploy are now a small fraction of those available for development from other sources. The search for distinctive initiatives may be frustrating.

As far as they go, the case-studies of past success show that science by itself is not enough. The case of Costa Rica, one of whose heroes is Leonardo Mata, the microbiologist head of the Institute for Health Research at Costa Rica's university, illustrates several points about the complexity of improvement.

Infant mortality has fallen from roughly one in five live births before 1930 to one in fifty now, to begin with only slowly as a consequence of public education (and Mata says that it helps merely to wear shoes), latterly more quickly by the control of malaria (by the 1960s), with the pursuit of the technology of oral rehydration therapy (for the prevention of infantile diarrhoea), arrangements for the collection and pooling of nursing mothers' milk and techniques of immunization.

The "silent army" of barefoot (but no doubt well-shod) doctors, originally opposed by more orthodox physicians, has done most of the field work, but the operation seems to have been helped on its way significantly by the social legislation left on the statute books by the government overthrown in the revolution of 1948. One accompaniment of this dramatic improvement in public health has been the decline of fertility (most rapidly in the late 1960s), but here again fertility rates had begun to fall, if slowly, before the technology of contraception had been made popular. Improvement is what would be called a multifactorial process, not simply a consequence of the availability of technology.

The same familiar truth emerges from the tale of India's most productive states of Punjab and Haryana, where the technology of the new high-yielding cereal varieties could be applied only after

socially adventurous land reforms had been carried through. And where, as in many parts of Africa, families hold to the view that large families are intrinsically desirable, the technology of contraception may have no discernible effect on fertility, but may merely allow women to space their children's births conveniently.

So what new technology can be exploited? The Rockefeller Foundation is flirting with biotechnology, and not surprisingly. (Excitement that malarial vaccines are now in trials is barely suppressed.) The tale at Bellagio was told by Howard A. Schneiderman of Monsanto, Charles Anzen of du Pont and Michael Sela, the immunologist past-president of the Weizmann Institute. That there will be pest-resistant crop plants and novel medicines for prophylaxis and therapy is beyond dispute. The unanswered questions concern the other criteria that will have to be satisfied before their benefits can be won in developing countries.

Two sets of questions are left hanging by all discussions of this kind. To visualize the way in which novel techniques could be used is not particularly difficult. But if, for example, the innovation is a plant with artificial advantages, what arrangements will there be to ensure that the seeds will find their way into the hands of farmers now distinguished by their lack of cash with which to purchase even the essential ingredients of their craft? And if experience in Costa Rica, echoed by that elsewhere, shows that public education is the foundation of public health, is it possible to hope for much improvement generally without alleviating the lack of teachers that marks the low-income countries?

There are two lessons to be learned from discussions of this kind, both of them all too familiar. First, development must be balanced in some fashion definable only by the social context in which it happens. And the underlying economic problems of the poorest countries feed on themselves. This does not mean that development is not possible but merely that it depends as much on good luck and imagination as on technology.

This contradiction is why the optimists talk of the chance that the rules may change, perhaps by the new agreement on trade two or three years from now. The other side of the coin is the waywardness of the governments whose people are the most in need of help. Technology is at least neutral.

John Maddox

AIDS research

Human and monkey virus puzzles

Peter Newmark

ETHICS and politics vied with virology and epidemiology for attention at the third International Conference on AIDS*. And PWAs (people with AIDS, as those who spoke at the meeting refer to themselves) were as much part of the meeting as HIVs, the human immunodeficiency viruses that cause AIDS. As for the science, although few real surprises emerged, the meeting provided the opportunity to assess what has been learnt in the past year. Here is the briefest of progress reports, with attributions only where identification of a single speaker is appropriate.

DNA sequencing of two isolates of HIV-2, the virus type that predominates in West Africa, has shown them to have over 80 per cent identical but to share only about 40 per cent of their sequences with HIV-1. A much closer relationship is that between HIV-2 and SIV (or STLV-III), the simian immunodeficiency virus, isolated from rhesus macaques and now sequenced. The similarity approaches identity in the case of the HIV-2 isolate called HTLV-IV and the SIV from African green monkeys, but many would interpret that as being a case of mistaken identity rather than a genuine result.

Consensus

In any case, the new consensus is that whereas SIV may have been the recent progenitor of HIV-2, the origin of HIV-1 is a puzzle. Either it has a direct progenitor of monkeys that has yet to be found, or it has evolved from HIV-1 in humans, in which case the viruses must have been in humans for longer than it seems, perhaps confined to a small population in which AIDS went unrecognized.

Any figures on the percentage relationship between different types of immunodeficiency virus have to be taken as first approximations because variations in sequence between different isolates of one type can be considerable. Variants can arise after infection; as many as 14 variants have been identified in one patient at one time, and there is some indication that they differ in biological activity. Variation is mostly by small mutations but there is a suspicion, based on lentivirus research, that recombination may also be possible (Simon Wain-Hobson, Pasteur Institute).

There is also some evidence to suggest that pathogenicity can evolve, at least in experimental animals; a virus (yet to be positively identified as SIV) from a pig-tailed macaque that died of an AIDS-like disease 14 months after being experi-

mentally infected with SIV from sooty mangabey monkeys, killed other pig-tailed macaques and sooty mangabeys within less than two weeks (Harold McClure, Yerkes Primate Research Center). The pathogenicity of HIV-2 remains a matter of controversy. Whereas Luc Montagnier's group (Pasteur Institute) has isolated HIV-2 from about 20 West Africans with typical AIDS and in most cases without any trace of a coinfection with HIV-1, Myron Essex and his colleagues have yet to find any evidence of pathogenicity associated with HIV-2 infections in their surveys and follow-ups of west Africans.

Two apparent relatives of HIV/SIV have emerged elsewhere in the animal kingdom. One causes an AIDS-like disease in cats, and was expected to be a variant of feline leukaemia virus. But by several criteria it is much more like HIV/SIV (Niels Pedersen, University of California, Davis). The other is from cows in which it causes lymphadenopathy (Matthew Gonda, Frederick Cancer Research Facility, Maryland). The cat virus has not yet been sequenced but a partial sequence of the cow virus shows that whereas it is firmly within the lentivirus family, it is probably more closely related to equine infectious anaemia virus than to HIV or SIV.

Experimental gene constructions and deletions have helped to sort out the functions of some of the genes of the HIVs, which have several more genes than most retroviruses or lentiviruses. The confusion over *tat* gene function has been more or less resolved with evidence that the protein it produces can control other genes either transcriptionally or post-transcriptionally, depending on circumstances. The *art* or *trs* gene product is also a positive regulator but only of structural proteins, of which the product of the 3'-ORF (open reading frame) gene may be a negative regulator. Deletion of the *sor* or *A* gene results in mutants with a much reduced ability to infect cells.

Intensive scrutiny of the *env* gene has led to the tentative identification of domains within the envelope protein that are involved in binding to the T4 molecule that acts as the virus receptor on T lymphocytes and, probably, brain cells. Segments of the protein that are recognized by antibodies and by helper T cells have also been identified. The discovery of a suppressible stop codon within the *env* gene of an SIV (James Mullins, Harvard School of Public Health), and its subsequent recognition in some HIV-2 clones, suggests that either full-length or

truncated forms of the envelope protein can be made and that their biological activities may differ. This can and will be put to the test.

There is increasing evidence of the presence of antibodies that neutralize HIV in the blood of infected people but the extent to which they will neutralize viruses other than those against which they have been raised remains an area of research, with importance for the attempt to produce broadly protective vaccines. Evidence that there is a T-cell mediated response against HIVs in addition to the B-cell mediated antibody response is mounting, with consistent reports of the presence of cytotoxic T cells directed against the viral envelope protein in infected but asymptomatic individuals.

Immunology and vaccines

Still the only human vaccine tests are those of Daniel Zagury (Universite Pierre and Marie Curie, Paris). He reported that 12 uninfected individuals including himself have now been immunized with a recombinant vaccinia virus that contains the full HIV-I envelope gene and boosted in one of several ways. The most successful boost consisted of the individual's own cells infected with the recombinant vaccinia virus and then inactivated. This produced broadly cross-reactive neutralizing antibodies. More practical boosters are being sought. Ten Africans with AIDS are involved in a trial of post-infection immunization (a procedure advocated by Jonas Salk, see page 473) using their own infected and inactivated cells but Zagury says it is too early to present any results.

Although trials of vaccinia-virus based vaccines in chimpanzees have had disappointing results, an application from Genetic Systems to proceed with human trials is being considered by the US Food and Drugs Administration. A second application is for a trial of a 30 amino-acid synthetic peptide analogue of the p17 product of HIV *gag* gene (Allan Goldstein, George Washington School of Health Sciences, Washington). For optimal presentation of the viral antigens in a vaccine to the immune system, William Jarrett (Glasgow University) advocated the incorporation of the antigens into ISCOMS. This technology has resulted in a fully protective vaccine against feline leukaemia virus and HIV ISCOMS are both good immunogens and non-toxic in gibbons and apes.

No vaccine will be widely available for at least five years. The only drug that is convincingly effective in treating AIDS, AZT (3'-azido-3'-deoxythymidine), is highly toxic. There are now more than 51,000 cases of AIDS in the world and with an estimated 5-10 million people infected, there will be 0.5-3 million cases in the next 5 years. □

Peter Newmark is Deputy Editor of Nature

* III International Conference on AIDS, Washington DC, 1-5 June 1987.

Vascular physiology

The end of the quest?

Paul M. Vanhoutte

ENDOTHELIAL cells react to mechanical force as well as to various neurohumoral mediators by releasing vasodilator and/or vasoconstrictor substances and as such contribute to the control of local vascular diameter. One of the most powerful substances released by the endothelium, endothelium-derived relaxing factor, has now been identified as nitric oxide. This finding, reported by Palmer, Ferridge and Moncada on page 524 of this issue¹, is the climax of one of the most exciting sagas in vascular physiology and pharmacology.

When Furchgott first discovered² that endothelial cells are obligatory in the relaxations evoked by acetylcholine in isolated arteries, few at first believed this to be anything but another *in vitro* artefact of no relevance to the normal behaviour of the blood vessel wall. Yet in the past six or seven years, an impressive number of experiments on isolated arteries from lower vertebrates, mammals and people have confirmed that indeed the presence of endothelial cells is a prerequisite to evoke relaxation with acetylcholine, with other neurohumoral substances, hormones and autacoids (Fig. 1) and with increases in shear stress. The findings on isolated blood vessels have now been confirmed in the intact organism³⁻⁶, and it is apparent that dysfunction of endothelial cells in controlling the underlying smooth muscle is an early sign of vascular disease such as atherosclerosis, hypertension and cerebral or coronary vasospasm⁸.

The obvious question is how does the

endothelium control the degree of contraction of the underlying smooth muscle? There are clearly two possibilities. First, cell-to-cell connections could pass the signal on from the endothelium to the smooth muscle; this possibility still has not been excluded at the microcirculatory level, where there are intimate connections (gap junctions) between endothelial and smooth muscle cells, or even in large arteries where fenestrae in the elastic laminae allow direct contacts between the two types of cells. The second explanation is that the endothelial cells, when stimulated by acetylcholine (or other endothelium-dependent relaxing agents), release an inhibitory substance(s). Bioassay studies using isolated blood vessels as donors of endothelial factors demonstrate beyond doubt that this is the case^{2,7,8}; further, it is obvious that endothelial cells in culture maintain the ability to produce the relaxing substances^{1,9}.

The identity of the substance(s) released by the endothelium remained elusive for several years, and Furchgott coined the term endothelium-derived relaxing factor to describe it. The term was immediately abbreviated to EDRF, giving everybody the impression of knowing what they were talking about. It was soon discovered that EDRF causes stimulation of the soluble guanylate cyclase of vascular smooth muscles, and that the resulting increased levels of cyclic GMP were probably responsible for the sustained endothelium-dependent relaxa-

tions caused by acetylcholine (and other endothelium-dependent relaxants)³⁻⁵; in certain vascular smooth muscle, EDRF also causes transient hyperpolarization, presumably through activation of sodium-potassium pumping (Fig. 2).

Once it was established that the endothelial cells could produce a potent relaxing substance, the quest for its identity started. Prostacyclin and other vasodilator prostanoids were easily excluded. A first hypothesis, that products of the lipooxygenase pathway for the metabolism of arachidonic acid were involved, did not stand the test of time. Subsequent studies demonstrated that EDRF is a very labile substance, with an extremely short half-life, which is most readily degraded by superoxide anions⁷⁻¹⁰. Last year, Furchgott¹¹ proposed that EDRF may be nothing else than nitric oxide (NO), which would explain the similarity of its action on vascular smooth muscle with that of synthetic nitrovasodilators. This interpretation has been reinforced considerably by the report of Palmer *et al.* in this issue¹, showing that endothelial cells in culture release NO, and that NO released from these cells is indistinguishable from EDRF in biological activity, stability and susceptibility to inhibiting and potentiating agents. Thus, NO and the EDRF that activates guanylate cyclase in vascular smooth muscle may be identical (Fig. 2).

Is the demonstration that NO is released by endothelial cells the end of the quest for the identity of EDRF? Probably not: several pharmacological observations strongly suggest that there is more than one EDRF (Fig. 2). Among the possible mediators are products of arachidonic acid through the P-450 cytochrome pathway or, more likely, ammonia (NH₃). The latter is an attractive candidate

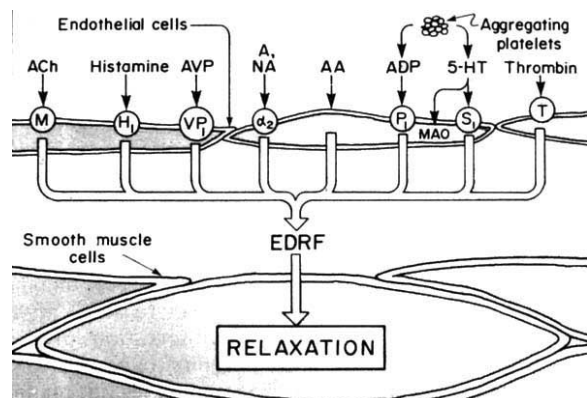


Fig. 1 Formation of a vasoactive substance(s) by the vascular endothelium: various substances may activate specific endothelial receptors (circles) and induce the formation of EDRF(s) which in turn causes relaxation of arterial vessels. A, adrenaline; ACh, acetylcholine; α₂, alpha-adrenoceptor; AA, arachidonic acid; AVP, arginine vasopressin; H₁, histaminergic receptors; 5-HT, 5-hydroxytryptamine (serotonin); MAO, monoamine oxidase; M, muscarinic receptors; NA, noradrenaline; VP₁, vasopressinergic receptors; P₁, purinergic receptors; S₁, serotonergic receptors; T, thrombin receptors.

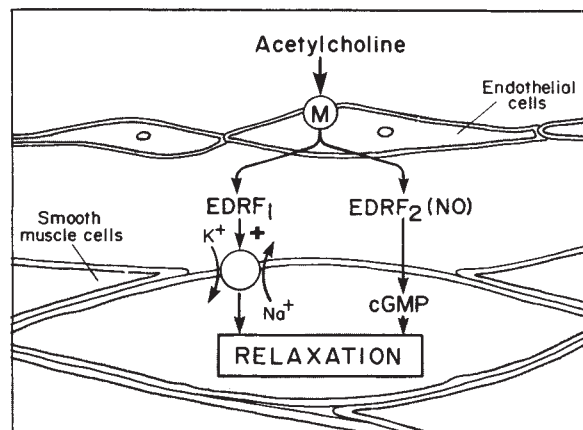


Fig. 2 In isolated blood vessels, acetylcholine causes relaxation by acting on muscarinic receptors (M) of endothelial cells, which in turn secrete one or more EDRF, causing inhibition of the contractile process in vascular smooth muscle by transient hyperpolarization followed by activation of guanylate cyclase to increase the production of cyclic GMP (cGMP). The work of Palmer, Ferridge and Moncada reported on page 524 suggests that the EDRF causing activation of guanylate cyclase is NO.

because endothelial cells possess several metabolic pathways that can lead to its formation and because it is a potent relaxant of vascular smooth muscle. If NH_2 does turn out to be an EDRF, it is intriguing that the endothelial cells have evolved to use very simple nitrogen derivatives (NO , NH_2) to signal a reduction in function to the underlying smooth muscle. This would probably be disappointing to modern pharmacologists eager to discover complex peptidergic transmissions between differentiated cells, but to me this result emphasizes the ancestral character of endothelium-dependent responses⁶. □

1. Palmer, R.M.J., Ferridge, A.G. & Moncada, S. *Nature* **327**, 524–526 (1987).
2. Furchgott, R.F. & Zawadzki, J.V. *Nature* **288**, 373–376 (1980).
3. Furchgott, R.F. *Circ. Res.* **53**, 557–573 (1983).
4. Peach, M.J. *et al.* *Hypertension* **7**, 194–1100 (1985).
5. Vanhoutte, P.M. *et al.* *Rev. Physiol.* **48**, 307–320 (1986).
6. Vanhoutte, P.M. *News Pharmac. Sci.* **2**, 18–22 (1987).
7. Griffith, T.M. *et al.* *Nature* **308**, 645–647 (1984).
8. Rubanyi, G.M., Lorenz, R.R. & Vanhoutte, P.M. *Am. J. Physiol.* **249**, H95–H101 (1985).
9. Gryglewski, R.J., Palmer, R.M.J. & Moncada, S. *Nature* **320**, 454–456 (1986).
10. Rubanyi, G.M. & Vanhoutte, P.M. *Am. J. Physiol.* **250**, H815–H821.
11. Furchgott, R.F. in *Mechanisms of Vasodilatation* (ed. Vanhoutte, P.M.) Ch. 1 (Raven, New York, in the press).

Paul M. Vanhoutte is in the Department of Physiology and Biophysics, Mayo Clinic, Rochester, Minnesota 55905, USA.

Planetary science

Measuring the Io plasma torus

Fran Bagenal

RATHER than twiddling their thumbs while waiting for delayed missions, planetary scientists have been busy making the most of the extensive datasets already available and looking for ways to apply remote sensing to planetary problems. Although local plasma measurements obtained on planetary fly-bys have generated a wealth of detailed information about plasma properties, they can only provide a snapshot of the magnetosphere. To understand the plasma processes it is often necessary to monitor temporal and spatial variability with remote sensing techniques. D. Jones, for example, on page 492 of this issue¹, draws our attention to the outer boundary of the plasma torus generated by Jupiter's innermost large satellite, Io. Radio emissions observed there for more than a month before and after each Voyager fly-by of Jupiter provide measurements of the radial gradient in plasma density.

The nature of the outer boundary of the torus is important in understanding how the plasma that is generated by Io (roughly at a rate of 1 tonne per second of sulphur and oxygen ions) subsequently spreads throughout the jovian magnetosphere and, ultimately, how the magnetosphere responds to changes in volcanic activity on Io. The initial analyses of the Voyager plasma data suggested that the density of thermal (<100 eV) plasma reaches a maximum value of more than 2,000 electrons cm^{-3} near Io's orbit at 6 Jupiter radii (R_J) and drops sharply between 7 and 8 R_J . This outer boundary of the torus is the same region where the number of energetic (10^3 – 10^6 eV) radiation-belt particles sharply increases with radial distance. The initial interpretation of these observations was that the radiation-belt particles have diffused inwards from a source in the outer magnetosphere and are scattered by the cooler, more dense, iogenic plasma (or accompanying plasma waves).

The scattered particles are diverted along the magnetic field into the upper atmosphere of Jupiter where they could stimulate the intense auroral emission observed in Jupiter's polar regions. The resulting radial gradient in particle pressure was thought to hinder the centrifugally driven outward diffusion of torus plasma, impounding the iogenic plasma and producing the sharp outer boundary of the torus.

In the past couple of years considerable effort has been put into verifying this picture, discussed at a symposium on the magnetospheres of the outer planets held at the University of Iowa last September. Although some people have concentrated on theoretical studies of the mechanism of plasma diffusion, others have been improving the observational constraints². Summers and Siscoe³ conclude that the pressure gradient of the energetic particles measured by the Voyager instrumentation is not sufficient to impound effectively the torus plasma. They calculate that about twice as many keV radiation-belt particles are needed if the thermal-particle density gradient is as steep as first reported by the Voyager experimenters. Summers, Thorne and Mei⁴ find they cannot generate a steep slope of outwardly diffusing thermal plasma even if the modification of ionospheric conductivity by precipitating particles is included.

Unfortunately the gap between the energy ranges covered by Voyager's plasma and particle instrumentation is in this key intermediate energy range and we will have to wait for the Galileo or Ulysses results for clarification. The Voyager plasma-study team, however, reported at Iowa that further analysis of the Voyager data suggests that the density of keV ions in the torus could have been underestimated and the gradient in the iogenic plasma may not be as steep as initially reported.

In a recent letter to *Nature* Jones⁵ used radio emissions to determine radial profiles of plasma density at the magnetic equator, which he found to be much shallower than the initial Voyager results. He has successfully applied his mechanism for linear conversion of electrostatic (upper hybrid) waves to electromagnetic (radio) emission to the magnetospheres of Jupiter, Saturn and the Earth. His mechanism requires that in the source region, the gradient of the plasma density is perpendicular to the magnetic field and produces a radio beam with characteristics given by the plasma density and magnetic field in the source region. Assuming the source of jovian broadband-kilometric emission to be at the magnetic equator, Jones used the geometry of a magnetic-field model and the Voyager trajectory to derive the radial distance for a particular emission frequency and corresponding plasma density. In his paper in this issue¹, Jones now applies the same emission mechanism to the narrowband-kilometric emission which he believes is produced at higher latitudes ($> 8^\circ$) on the flanks of the torus. At these latitudes there are no direct plasma measurements, but Jones's preliminary results are compatible with density models based on Voyager data.

At present the laborious nature of the technique means it is not particularly accurate for determining plasma density. But it could be a means of testing the diffusion hypothesis without waiting for the Galileo or Ulysses missions. The technique also promises to be a powerful tool for investigating plasma processes by monitoring time-variability of the outer torus boundary.

Time-variable phenomena in the jovian system will be discussed this summer in Flagstaff, Arizona. The International Jupiter Watch, inspired by the successes of the International Halley Watch, has been set up to co-ordinate observations of the planet and its satellites and magnetosphere. It was evident at the Iowa meeting that there exists a wealth of particle, field and ultraviolet data obtained by Voyager at Jupiter that have rather been neglected because of the Saturn and Uranus encounters, to which experimenters are keen to return. Moreover, with planetary missions becoming scarce, magnetospheric physicists are increasingly appreciating the value of observations of plasma emissions from radio to X-rays for explaining plasma processes in planetary magnetospheres. □

1. Jones, D. *Nature* **327**, 492–495 (1987).
2. Richardson, J.D. & McNutt, R.L. Jr *Geophys. Res. Lett.* **14**, 64–67 (1987).
3. Summers, D. & Siscoe, G.L. *J. geophys. Res.* **90**, 2665–2671 (1985).
4. Summers, D., Thorne, R.M. & Mei, Y. *Planet. Space Sci.* (in the press).
5. Jones, D. *Nature* **324**, 40–42 (1986).

Fran Bagenal is at the Blackett Laboratory, Imperial College, Prince Consort Road, London SW7 2BZ, UK.

Evolutionary ecology

El Niño and Darwin's finches

Jon Seger

LIKE astronomers, evolutionary biologists can seldom make direct experimental tests of their central theories, so they rely instead on natural perturbations that can later be interpreted as if they had been planned. Astronomy now has a supernova, and evolution has the aftermath of the recent El Niño – Southern Oscillation event, which briefly but dramatically altered weather patterns in many parts of the western hemisphere. The Galápagos islands experienced eight months of record-breaking rainfall between November 1982 and July 1983. Many plant species responded with equally record-breaking levels of seed production. Several species of Darwin's finches bred continuously throughout this extraordinary eight-month wet season, and some of the offspring born early in the season had matured and bred successfully before it was over^{1,2}. On page 511 of this issue³, Gibbs and Grant show that during the next two years (1984–85) there was strong natural selection for reduced body size in one population of the seed-eating ground finch *Geospiza fortis*.

Rainfall on the Galápagos is always highly variable and, as a consequence, so is seed production. During the dry years of 1976–77 and 1981–82, Grant and co-workers^{2,4,5} found that there was selection for increased body size (and in particular, for increased bill depth) within this same population of *G. fortis* (see figure), and they traced the ecological cause to a changing distribution of seed sizes. Small, relatively soft seeds are eaten by both small and large birds. Thus, as the seed supply shrinks during a period of low rainfall, it becomes dominated by large, relatively hard seeds that are more easily cracked and eaten by large birds than by small ones. The two years following El Niño were very dry, but small seeds were more abundant than they had been during previous dry years, owing to the huge crop produced in 1983 (see cover picture of this issue). This appears to be the reason why Gibbs and Grant found a pattern of selection essentially opposite to what would otherwise be expected during a drought. One implication is that small birds tend to use small seeds more effectively than do large ones¹, which in turn implies that selection on bill and body sizes often includes a frequency-dependent component. It should be emphasized that these studies concern only differential survival, not differential reproduction among the survivors, which also occurs in this population^{6,7}. Nonetheless, evolutionary responses of the expected signs

and magnitudes have been observed².

Many previous studies have demonstrated selection in nature, but as Endler⁸ emphasizes in a recent review, there are few cases in which the ecological causes of differential survival or reproduction have been clearly identified. Endler's list of notable exceptions includes studies on two genera of Lepidoptera, one fish, three plants and *Geospiza*. In only one of these cases has selection been shown to have acted in opposite directions within the same population at different times, and in that case (industrial melanism in the moth *Biston betularia*) the ecological cause was of human origin. The new work by Gibbs and Grant³ shows that major



Two individuals of *G. fortis* on Isla Daphne Major, showing variation in bill shape.

natural episodes of selection in opposite directions can follow each other at short intervals.

The six morphological characters studied by Gibbs and Grant³ seem to have served mainly as proxies for overall body size; none stands out as a distinct target of selection. But during earlier droughts, when selection generally favoured large size, differences in the intensity and even in the direction of selection on these six genetically correlated characters had been detected. If such strong and constantly changing patterns of selection should turn out to be the rule for many species (as they seem to have been for *G. fortis* on Isla Daphne Major during the past decade), then we are probably far from understanding the processes that maintain heritable quantitative variation in natural populations^{9,10}. □

1. Gibbs, H.L. *et al.* *Auk* **101**, 872–874 (1984).
2. Grant, P.R. *Ecology and Evolution of Darwin's Finches* (Princeton University Press, 1986).
3. Gibbs, H.L. & Grant, P.R. *Nature* **327**, 511–513 (1987).
4. Boag, P.T. & Grant, P.R. *Science* **214**, 82–85 (1981).
5. Price, T.D. *et al.* *Nature* **309**, 787–789 (1984).
6. Price, T.D. *Evolution* **38**, 327–341 (1984).
7. Price, T.D. *Am. Nat.* **123**, 500–518 (1984).
8. Endler, J.A. *Natural Selection in the Wild* (Princeton University Press, 1986).
9. Turelli, M. *Genetics* **111**, 165–195 (1985).
10. Turelli, M. in *Evolutionary Processes and Theory* (eds Karlin, S. & Nevo, E.) 607–628 (Academic, Orlando, 1986).

Jon Seger is at the Department of Biology, University of Utah, Salt Lake City, Utah 84112, USA.

Cellular immunology

Interleukin-2 receptor proteins

Ian Trowbridge

THE proliferative response of mature T lymphocytes to stimulation by antigens is regulated by the lymphokine interleukin-2 (IL-2). The production of IL-2 by a subset of helper T cells and the induction of specific IL-2 receptors on T cells occur as a consequence of activation by the antigen. Activated T cells express two classes of IL-2 receptor that differ in their affinity for IL-2. About 10 per cent of the receptors are high affinity (dissociation constant (K_d) ~ 10 pM) and appear to mediate the physiological response of T cells to IL-2; the remaining receptors bind IL-2 with much lower affinity ($K_d \sim 10$ nM). The structural basis for these two classes of receptors has remained unclear. A cell-surface glycoprotein, p55, of relative molecular mass 55,000 (55K), first identified on activated human T cells by the monoclonal antibody anti-Tac, binds IL-2 and has been implicated as a component of both classes of receptor. Recent results from several laboratories^{1–4}, including the report by Dukovich *et al.* on page 518 of this issue⁵, provide strong evidence that there is a second IL-2 binding protein that could interact with p55 to

form the high-affinity IL-2 receptor.

The search for other structural components of the IL-2 receptor was prompted by several observations that challenged the view that p55 was the only molecule involved in binding IL-2 and triggering a physiological response. Whereas transfection of nonlymphoid mouse cells with complementary DNAs encoding human p55 generated only low-affinity receptors⁶, similar transfectants of two mouse T-cell lines displayed both high- and low-affinity human IL-2 receptors^{7,8}. These results suggested that T cells express a cell-specific component that modulates the binding affinity of p55. Further, the deduced structure of p55 predicts that only 13 amino acids form the cytoplasmic domain, raising questions as to the mechanism of signal transduction^{9–11}. Finally, some lymphoid cells, including natural killer cells and precursors of lymphokine-activated killer cells, apparently do not express p55 but nevertheless bind and respond to IL-2.

The approach taken by Dukovich *et al.* in this issue⁵ and by the other groups^{1–4} to characterize the molecules involved in

IL-2 binding was to bind 125 I-labelled IL-2 to cells under high- and low-affinity conditions, and then to crosslink using the bifunctional reagent disuccinimidyl suberate. These studies have led to two major conclusions that can be stated with different degrees of certainty. First, despite the intrinsic limitations of these crosslinking experiments, there is compelling evidence for a novel IL-2 binding protein, p70. The most clear-cut data come from the study of two cell lines, the gibbon ape cell line MLA-144 (refs 2,5), and a clone, 2C2, of the human leukaemic cell line YT (refs 3,4). Both cell lines bind IL-2 with intermediate affinity ($K_d \sim 1$ nM), but do not express p55. They neither contain detectable p55 messenger RNA nor react with the anti-Tac antibody. In the case of MLA-144 cells, the lack of reactivity with anti-Tac antibody does not result from the species difference as the antibody reacts strongly with normal activated gibbon ape T cells. Crosslinking of IL-2 bound to MLA-144 or YT-2C2 cells gives a single 85–90K labelled product, consistent with one molecule of IL-2 (15K) binding to a cell membrane receptor of ~ 70 K. Dukovich *et al.*⁵ detected a similar species on purified large granular lymphocytes; on a B-cell line, SKW6.4B, that secretes immunoglobulin in response to IL-2; and, surprisingly, on at least some unstimulated CD3⁺ T cells.

Second, it seems likely that high-affinity IL-2 receptors are composed both of p55 and of p70. Under low- and high-affinity conditions, there are two main labelled species on normal activated T cells and HUT102B2 cells that also express high- and low-affinity receptors. One is a cross-linked product (~ 70 K) that reacts with anti-p55 antibodies; the other does not react with anti-p55 antibodies and is similar in size to the crosslinked product obtained from MLA-144 or YT cells. The relative amount of the heavier band (~ 90 K) is reduced under low-affinity conditions. These results have been interpreted to indicate that this band is derived from high-affinity receptors and represents the crosslinking of IL-2 to p70. Dukovich *et al.* further show that the introduction of complementary DNA encoding p55 into MLA-144 cells by protoplast fusion generates high-affinity binding sites. Similarly, other work^{3,4} indicates that YT cells induced by various treatments to display high-affinity IL-2 receptors concomitantly express both p55 and p70.

Thus, the available data are consistent with the idea that high-affinity IL-2 receptors are composed of at least two subunits, each of which is independently capable of binding IL-2 with much lower affinity. There is evidence that each subunit interacts with a different region of the IL-2 molecule, supporting the notion that both directly contribute to the binding site of the high-affinity receptor¹. But as yet

there is no direct evidence for a ternary complex involving p55, p70 and IL-2. Further, without antibodies specific for p70, the relationship between the cross-linked 90K bands from activated T cells and MLA-144 cells can only be inferred. There are precedents both for cell-surface receptors that exhibit two classes of binding sites of different affinities and for multi-chain receptors in which two polypeptides contribute to the ligand binding site. There are high- and low-affinity forms of receptors for epidermal growth factor and nerve growth factor. The low-affinity form of nerve growth factor has recently been cloned^{12,13} and it has been suggested that the high-affinity form contains this polypeptide together with an additional subunit¹³. In the B-cell antigen receptor, immunoglobulin heavy and light chains form part of the ligand binding site. The putative high-affinity IL-2 receptor is unusual in that each subunit independently has measureable binding affinity.

The conclusion that high-affinity IL-2

receptors consist of at least two subunits has important implications: structural studies of p70 may help to elucidate signal transduction by the high-affinity receptor. The role of IL-2 during thymic differentiation is still obscure, but the discovery of the second IL-2 cell-surface binding protein may also open up new avenues of investigation in this field. □

1. Sharon, M., Klausner, R.D., Cullen, B.R., Chizzonite, R. & Leonard, W.J. *Science* **234**, 859–863 (1986).
2. Tsudo, M., Kozak, R.W., Goldman, C.K. & Waldmann, T.A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9694–9698 (1986).
3. Teshigawara, K., Wang, H.-M., Kato, K. & Smith, K.A. *J. exp. Med.* **165**, 223–238 (1987).
4. Robb, R.J., Rusk, C.M., Yodoi, J. & Greene, W.C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2002–2006 (1987).
5. Dukovich, M. *et al.* *Nature* **327**, 518–521 (1987).
6. Greene, W.C. *et al.* *J. exp. med.* **162**, 363–368 (1985).
7. Hatakeyama, M. *et al.* *Nature* **318**, 467–470 (1985).
8. Kondo, S. *et al.* *Nature* **320**, 75–77 (1986).
9. Leonard, W.J. *et al.* *Nature* **311**, 626–631 (1984).
10. Nikaïdo, T. *et al.* *Nature* **311**, 631–635 (1984).
11. Cosman, D. *et al.* *Nature* **312**, 768–771 (1984).
12. Johnson, D. *et al.* *Cell* **47**, 545–554 (1986).
13. Radeke, M.J., Misko, T.P., Hsu, C., Herzenberg, L.A. & Shooter, E.M. *Nature* **325**, 593–597 (1987).

Ian Trowbridge is at the Salk Institute, PO Box 85800, San Diego, California 92138-9216, USA.

Photosynthesis

A quantum jump for chemistry

John M. Warman

CHEMISTS have known for many years the individual molecular components essential to the working of nature's photosynthetic machinery. Despite this nobody has managed to put the pieces back together to recreate the primary natural function — that of a solar-powered, electrolytic battery. Not for the first time, chemistry seemed to be confronted by a vital force — a secret known only to nature herself. A new approach to chemical synthesis, exemplified by the report of Joran, Dervan and their colleagues on page 508 of this issue¹, has, however, led to a breakthrough in this reconstructive problem.

Many secrets of photosynthesis have been revealed in the recent, detailed X-ray structural analysis of a bacterial photosynthetic reaction centre by Deisenhofer *et al.*² (discussed in a previous News and Views article³). In particular Deisenhofer *et al.* determined the precise spatial relationships between the molecular entities responsible for the conversion of absorbed photons into a voltage difference across the membrane. It is this voltage that drives the chemical stages of photosynthesis. Even before the work of Deisenhofer *et al.* many aspects of the photovoltaic stage had been deduced. It was known, for example, that within a few picoseconds of a photon reaching the reaction centre an electron is transferred from a sandwich-like pair of bacteriochlorophyll molecules to a nearby bacteriopheophytin. This is followed within

200 ps by a second electron transfer from the pheophytin to a quinone molecule resulting in virtual stabilization of the charge separation, allowing subsequent processes to occur at a more leisurely pace (see the recent discussion by M.C.W. Evans in News and Views⁴).

What is still puzzling is that the molecules involved in these primary electron-transfer steps seem to be rather widely separated, by distances of approximately 10 Å. Furthermore, there is no evidence that they are connected by molecular 'wires' for electrons to flow through. The active components are rigidly embedded in a non-polar region of the transmembrane protein matrix. Charge separation, therefore, must occur by electron tunnelling between the components. Quantum-mechanical tunnelling enables electrons to move across relatively large, atomic distances without ever being in between. Somehow, the idea that primitive biological systems evolving billions of years ago would have incorporated into their metabolisms quantum-mechanical devices only thought of a few decades ago is, even for many scientists, somewhat hard to take.

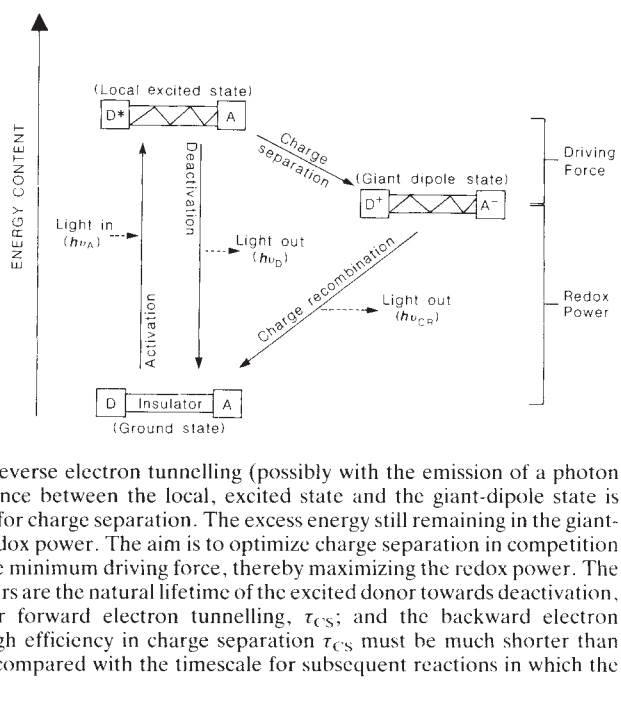
To elucidate the electron-tunnelling process, compounds are synthesized with barriers to electron transfer deliberately built in — an approach that could be described as molecular engineering. In this spirit, the compound thus synthesized is considered as a device fulfilling a physico-chemical function: that of a rectifier for converting electromagnetic waves into a

d.c. voltage. The basic design involves a donor group with a low ionization potential, an acceptor with an affinity for electrons, and a hydrocarbon spacer or bridging unit that is (in the first instance) electrically insulating — an inert separator of the donor and acceptor. The spacer should be as rigid as possible to give a well-defined geometry and also to prevent the two ends coming together and short-circuiting the molecule after charge separation. Such compounds are called DIADs, for donor-insulator-acceptor devices. The essential difference between DIADs and old-fashioned, donor-acceptor compounds is illustrated in Fig. 1.

The synthesis of DIADs is now being performed in many laboratories. The recent contribution^{5,7} of Mike Paddon-Row, who has made a beautiful series of variable-length DIADs with complete distance and orientational rigidity, is undoubtedly an important pioneering step. The field is not lacking in romance, especially when the molecular components used closely resemble those found in nature, as is the case in the work of Joran *et al.*¹. The most renowned of such compounds are the three-component TRIAD assemblies of Moore *et al.*⁸ reported in *Nature* and discussed in an accompanying News and Views article⁹. These molecules were composed of carotene, porphyrin and quinone residues, a gorgeous molecular cacophany that received worldwide press attention.

Many parameters need to be investigated for long-distance electron transfer to be understood and controlled. The role of the driving force for charge separation

Fig. 2 Ergodynamics of a photoexcited donor-insulator-acceptor device (DIAD). Activation by absorption of a photon $h\nu_A$ by the donor chromophore results in local electronic excitation. This is followed by deactivation back to the ground state (possibly by emitting a photon $h\nu_D$), or by tunnelling of an electron through the insulating barrier to the acceptor resulting in complete charge separation. The giant dipole so formed eventually undergoes charge recombination by reverse electron tunnelling (possibly with the emission of a photon $h\nu_{CR}$). The energy difference between the local, excited state and the giant-dipole state is known as the driving force for charge separation. The excess energy still remaining in the giant-dipole state is called the redox power. The aim is to optimize charge separation in competition with deactivation, using the minimum driving force, thereby maximizing the redox power. The important kinetic parameters are the natural lifetime of the excited donor towards deactivation, τ_D ; the time required for forward electron tunnelling, τ_{CS} ; and the backward electron transfer time, τ_{CR} . For high efficiency in charge separation τ_{CS} must be much shorter than τ_D , and τ_{CR} must be long compared with the timescale for subsequent reactions in which the redox power is used.



(Fig. 2) is clearly of prime importance. Without a driving force, electron transfer could not compete with deactivating processes after excitation by light. But, as can be seen from Fig. 2, the energy used in charge separation is lost, unavailable for further chemical or physical work. So it is important to know the minimum driving force necessary to compete efficiently with deactivation.

Herein lies the importance of the work of Joran *et al.*¹. By tinkering with the sub-

stituents of the quinone acceptor in their basic DIAD structure they are able to vary the driving force for electron transfer from the zinc-porphyrin donor by 0.6 eV without appreciably changing the separation of 10 Å between donor and acceptor. They find that a marked increase in the rate of charge separation, by more than an order of magnitude, occurs over the first 0.2 eV; above this the rate of charge separation is almost constant, the process taking about 100 ps as in the second step of photosynthesis. It is significant that for a given compound the rate of charge separation is almost independent of the solvent, be it non-polar benzene or highly polar acetonitrile. This has also been found^{5,6} for the forward electron-transfer step of Paddon-Row compounds. The eventual stabilization energy gained from specific ion-polar-solvent interactions in the giant dipole state apparently has little influence on the forward kinetics. Although, with hindsight, this should be expected, as solvent reorganization should occur after transfer, the result must be verified.

The insensitivity of charge-separation kinetics to solvent polarity is in stark contrast with the very strong solvent influence on charge recombination^{6,7,10}. The rate of long-distance charge recombination increases by two orders of magnitude when the solvent medium is changed from a non-polar saturated hydrocarbon to the more highly solvating ether dioxane⁶. This contradicts the intuitive feeling that ionic species are more stable in more polar media, but evidently is highly favourable for photosynthetic compounds working in a non-polar environment.

The study of specific medium effects on the highly exothermic recombination of

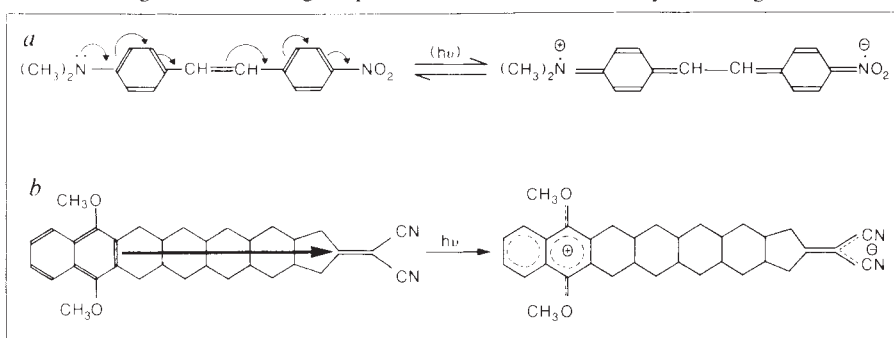


Fig. 1 Examples of classical and quantum-mechanical charge-transfer compounds: *a*, dimethylamino-nitrostilbene (DMANS), a classical donor-acceptor compound in which the electron-rich, donor nitrogen of the dimethylamino group, on the left, is separated from the electron-deficient, acceptor nitrogen of the nitro group on the right by a system of conjugated double bonds. As shown by the arrows, one of the lone pair of electrons on the amino-nitrogen (indicated by two dots) can move to the nitro group. On photoexcitation the dipole moment of DMANS increases from 7 to 24 Debye, considerably less than the 60 Debye corresponding to complete transfer of an electron. The conjugated hydrocarbon bridge conducts charge in both directions so that complete charge separation can never be attained for this type of compound. *b*, One of the Paddon-Row compounds^{5,7}. The donor, dimethoxynaphthalene, is separated from the acceptor, dicyanoethylene, by a rigid, non-conjugated paraffinic bridge (additional groups that give rigidity have been omitted for clarity). In the ground state there is no interaction between donor and acceptor. In ultraviolet light the donor becomes electronically excited and within about 100 ps an electron jumps from the excited donor to the acceptor, over a distance of almost 15 Å. Even in non-polar solvents the charge separation is complete, resulting in a giant dipole of almost 70 Debye. Such giant dipoles can apparently be stable for up to almost a microsecond — long enough to drive secondary, chemical or physical processes. This behaviour can be explained only by quantum-mechanical tunnelling of an electron through the intervening barrier (see text).

giant dipoles should be extremely interesting. The high-exothermicity, so-called inverted region of electron-transfer kinetics has been a source of continuous controversy for 30 years since Marcus¹¹ derived the relationship given as equation (1) in the paper of Joran *et al.*. There is growing theoretical interest in the question of what role, if any, is played by the intervening medium in the tunnelling process. Terms such as 'superexchange' and 'through-band interactions' are used in an attempt to describe this non-classical behaviour. This may be of particular relevance to the mechanism of photosynthesis as structural analysis² shows that an extra bacteriochlorophyll and two non-conjugated phytol side-chains lie along the path followed in the primary electron-transfer steps. This may not be just a coincidence.

The novel, rigid, donor-insulator-acceptor compounds now being synthesized are significant far beyond simply being humble imitations of the wonderfully complex machinery of natural photosynthetic compounds. With them it is possible to test general theories of long-distance electron-transfer and microscopic ion-solvent interactions far more rigorously than has previously been possible. Note that the insulating barriers in molecular DIADs are expected to be much better characterized in size and chemical composition than barriers prepared by physical means. The results will not be limited to systems of chemical interest but will be relevant to any process in which electron tunnelling is important.

Although practical applications are a long way off, one needs only ponder the use of DIADs as molecular-sized 'solar batteries' or 'digital switches' to realize their enormous inherent potential in energy storage, catalysis, optoelectronics and so on. Many of the additional physico-chemical tricks necessary to bring about these applications — the controlled growth of well-characterized surface films or the manufacture of nanometre-sized artificial membranes — have developed greatly in recent years. A fascinating future clearly lies ahead for a new generation of molecular engineers. □

- Joran, A.D. *et al.* *Nature* **327**, 508–511 (1987).
- Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. *J. molec. Biol.* **180**, 385–398 (1984).
- Barber, J. *Nature* **315**, 278–279 (1985).
- Evans, M.C.W. *Nature* **327**, 284–285 (1987).
- Verhoeven, J.W., Paddon-Row, M.N., Hush, N.S., Overing, H. & Heppener, M. *Pure appl. Chem.* **58**, 1285–1290 (1986).
- Paddon-Row, M.N. *et al.* *Abstr. 6th Int. Conf. Photochemical Conversion and Storage of Solar Energy, Paris*, A49, A50 (1986).
- Warman, J.M. *et al.* *Nature* **320**, 615–616 (1986).
- Moore, T.A. *et al.* *Nature* **307**, 630–632 (1984).
- Barber, J. *Nature* **307**, 596 (1984).
- Wasielewski, M.R., Niemczyk, M.P., Svec, W.A. & Pewitt, E.B. *J. Am. chem. Soc.* **107**, 1080–1082 (1985).
- Marcus, R.A. *J. chem. Phys.* **24**, 966–978 (1956).

John M. Warman is in the Radiation Chemistry Group, Interuniversity Reactor Institute, Mekelweg 15, 2629 JB Delft, The Netherlands.

Gene regulation

Specificity and flexibility

Miranda Robertson

MUCH recent thinking about eukaryotic gene regulation reflects two important simplifying assumptions drawn principally from observations on prokaryotes. First, regulatory proteins bind to DNA in its canonical right-handed helical conformation. Second, transcriptional activation is mediated by specific interactions between proteins bound close to the transcriptional initiation site or at distant sites (for a recent partisan view of the evidence, see ref. 1). The influence of these ideas was conspicuous at a recent meeting*, where it has to be said that their limitations also began to emerge. This is, however, because there is still much about eukaryotes that is not understood, rather than because of any serious threat to the validity of the fundamental insights derived from investigations on prokaryotes.

On the contrary, there seems to be a real possibility that the molecular mechanisms of embryogenesis in the fruitfly

Drosophila can be understood as a baroque evolutionary elaboration of those governing the life cycle of bacteriophage λ , and may be approachable by the same experimental route. This prospect, which is opened up by the experiments of Mark Krasnow (David Hogness's laboratory, Stanford University), can be properly appreciated only against some background in *Drosophila* developmental genetics and will be discussed in a later article, in the context of the account of molecular strategies of gene regulation that now follows.

Repressors and DNA flexion

The simplest device for transcriptional repression is a molecule that binds tightly to an operator and obstructs the access of transcriptional activators and RNA polymerases. An extraordinarily powerful approach to the structural basis of protein-DNA interactions is provided by the availability in bulk of specific DNA

*Institute of Genetics, Cologne, 3–6 March 1987.

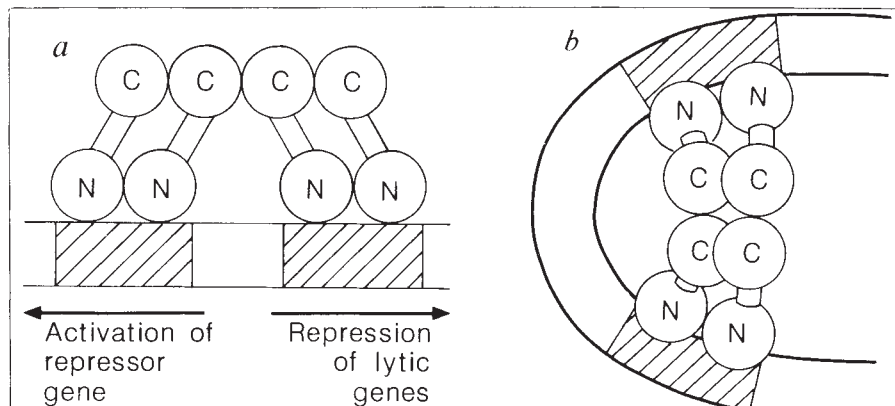


Fig. 1 *a*, Schematic diagram of two λ -repressor dimers cooperatively bound to adjacent operator sites (shaded); *b*, DNA looping consequent on cooperative binding of two repressor dimers to artificially separated operator sites. In *a*, the amino terminal of the repressor is shown binding specifically to the DNA; the carboxyl terminals mediate the protein-protein interactions responsible for dimerization and cooperativity. In a lysogenic bacterium, the λ genome is integrated into the bacterial DNA and replicates with it, without producing lytic phage particles. The lysogenic state is maintained by the cooperative binding of repressor dimers to adjacent operator sites, repressing the lytic genes to the right and activating their own gene to the left, maintaining repressor concentration by positive feedback and stabilizing the lysogenic state. The switch to the lytic state is induced by environmental stress which activates a protease that cleaves the repressor between the N-terminal and the C-terminal domains, eliminating cooperativity. This results in the dissociation of the repressor from the DNA so that further repressor synthesis ceases. At the same time the lytic genes to the right are released from repression and this allows the synthesis of a second regulatory protein, *cro*, that binds to the same operator sites, preventing further repressor binding and stabilizing the lytic state. In *b*, repressor is bound to artificially separated operators. Because repressor binds only on one side of the DNA, looping occurs only if the two operator sites are separated by an integral number of helical turns of the DNA. If the two sites are separated by a non-integral number of turns, the repressors will bind on opposite sides of the DNA and cooperative binding would require twisting as well as looping. Presumably because this makes excessive demands on DNA flexibility, cooperative binding does not occur in these circumstances. This can be detected in two ways. First, the affinity of the repressor molecules for their operators remains at the level expected of non-cooperative binding; and second, nuclease hypersensitivity consequent upon DNA looping does not appear if the two operators are a non-integral number of turns apart. Thus, DNase hypersensitivity shows a marked periodicity with distance between operator sites.

sequences and the protein domains that bind to them, and of techniques for introducing discrete mutations into either of the two interacting molecules. Analysis at this level has been achieved only for prokaryotic regulatory interactions and its most recent refinements reveal the need for some interesting and important modifications to the simplifying assumptions about regulatory interactions outlined above, as illustrated in the account by Stephen Harrison (Harvard University) of the structural basis of operator binding by the repressor protein of the lambdoid phage 434 (refs 2, 3).

The 434 repressor binds as a dimer to a bilaterally symmetrical operator site, the amino-terminal domains of the dimer having a classical helix-turn-helix DNA-binding motif in which helix three serves as the recognition helix and makes specific contacts with the bases in the major groove of the operator DNA. What the high-resolution structure of the repressor-operator cocrystal now reveals is that the correct positioning of the recognition helices requires gentle flexion and overwinding of the operator DNA, with concomitant narrowing of the minor groove and thus a slight but significant departure from the canonical B-DNA structure². Moreover, this distortion depends on the base composition of DNA that does not touch the protein: if the A-Ts at the centre of the operator are replaced with C-Gs, the binding affinity of the repressor is reduced 40-fold — presumably because flexion of the DNA is less easily accommodated by the substituted bases³.

This and other recent reports^{4,5} suggest that distortion of the DNA on specific protein binding may be quite general^{6,7}; in the case of the bacterial transcriptional activator protein CAP, the operator site is grossly bent, and Liu-Johnson *et al.* have speculated that the free energy of DNA bending may contribute significantly to the melting of DNA on polymerase binding⁸. At an estimated 2–3 kcal mol⁻¹, the free energy of flexion around the 434 repressor is not in the same league; but it may have equally important functional implications. Whether a phage multiplies and lyses an infected bacterium or integrates into its chromosome as a silent passenger (lysogen) is a decision that rests on differences of just this energy order in the binding affinities of two proteins (repressor and *cro*) for their operator sites, some of which may be occupied by either protein⁹. The central non-contacted bases in the operator sites vary considerably and may contribute significantly to the stability of the alternative states maintained by the differential binding of the two proteins (see Fig. 1).

Release from repression

The release of the lambdoid repressors from their binding sites and the concomit-

ant switch to the lytic cycle depend on a somewhat drastic process that eliminates cooperative binding of the repressor dimers to the DNA without, however, changing the structure of the DNA-binding site itself (see Fig. 1 legend). A very different strategy is embodied in the *trp* repressor, whose crystal structure in both the active repressor state⁹ and the inactive aporepressor state¹⁰ have been solved by Paul Sigler (Chicago University). The *trp* repressor of *Escherichia coli* controls the operon that encodes the enzymes required for tryptophan biosynthesis and in the absence of tryptophan binds DNA non-specifically and with low affinity. When tryptophan is abundant, two molecules bind to each molecule of the aporepressor which undergoes a conformational change enabling it to bind tightly and specifically to the *trp* operator, shutting off further tryptophan synthesis.

The *trp* repressor shares with 434 repressor the helix-turn-helix motif, two pairs of movable DNA-binding helices being disposed symmetrically on either side of a rigid central core. The part played by the bound tryptophan is to promote hydrophobic interactions that lock the recognition helices into the active position, in which they are thrust into the major groove of the DNA (see Fig. 2). When the tryptophans dissociate, the recognition helices fold inward towards the core, losing their fit to the DNA; and an alternative set of interactions forms in the tryptophan-binding pocket, stabilizing the aporepressor conformation.

Unlike Harrison, Sigler and collaborators have not yet solved the repressor-operator cocrystal structure and their analysis is based on the assumption that the repressor binds to idealized B-DNA. That this is probably not the case is suggested by the sensitivity of repressor binding affinity to mutations in positions not predicted to contribute to recognition.

The relative importance of specific recognition and conformational flexibility is brought into still sharper focus by recent analyses of transcriptional activation.

Protein flexibility

Transcriptional activation is more demanding than repression: activators must not only bind DNA but launch polymerases. The mechanism by which this is achieved is unknown, although there is evidence strongly suggesting that activators work by positioning RNA polymerases on the DNA rather than, for example, by changing DNA conformation so as to promote melting (Henri Buc, Pasteur Institute; see refs 8, 11).

In either case, it seems likely that activation is more easily disrupted than repression, and Ann Hochschild in Mark Ptashne's laboratory (Harvard University) has demonstrated this point in an extension of their experiment on DNA

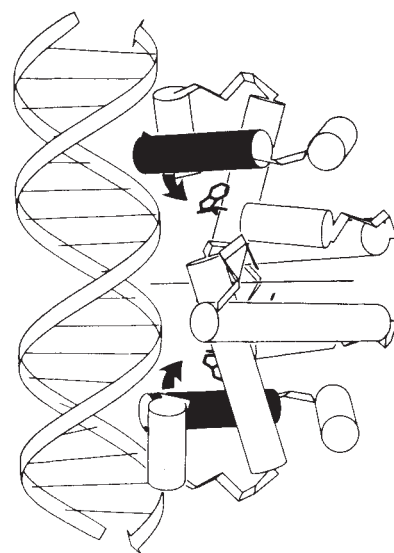


Fig. 2 The *trp* repressor. When the tryptophan molecules (hexagons) dissociate, the recognition helices (black) swing inwards (arrows) and no longer engage with the major groove of the DNA.

looping induced by the cooperative binding of two λ -repressor dimers to artificially separated operator sites. This experiment¹², with the seminal work of Schleif¹³, has become a paradigm for the study of cooperative binding and DNA looping and is explained in the legend to Fig. 1.

What it shows is that cooperative binding and *ipso facto* repression can survive the separation of two normally adjacent operator sites. But λ repressor is a transcriptional activator as well as a repressor (see Fig. 1), and transcriptional activation is abolished in these constructs. The basis for this sensitivity is unknown, but it is assumed that the topological distortion disrupts some specific protein-protein interaction required for polymerase binding. With the attempt to define such specific interactions in eukaryotic transcriptional activation, however, the simplified prokaryotic picture starts to dissolve and flexibility begins to look more important than specificity.

It has been neatly demonstrated more than once that the DNA-binding function of a eukaryotic transcriptional activator can be separated from the domain conferring the activation function: thus a yeast transcriptional activator will still work if it is attached to the DNA-binding domain of a bacterial protein, provided that the correct bacterial promoter sequence is inserted upstream of the yeast gene^{14,15}. The implication of these experiments is that while one domain of the regulatory protein confers specific interactions with the DNA, the other independently confers specific interactions with other proteins.

Recent attempts by Ian Hope and Kevin Struhl (Harvard University) to define the structural basis of such activating interactions have, however, produced

a series of surprising results. Using a bacterial DNA-binding domain linked to the yeast transcriptional activator GCN4, they showed that transcriptional activation survives the deletion of all but a central 19 amino acids within the GCN4 protein¹⁵. These amino acids are too few to have a stable folded structure and have no conspicuous homology to other transcription-activating sequences, with which they share only the general property of being relatively acidic; and more recent experiments reveal that other deletions that include these 19 amino acids will still leave a functional activation domain (Hope).

Similarly disconcerting results have emerged from experiments on a second yeast activator, GAL4 (Jun Ma in Ptashne's laboratory)¹⁶. In the most recent and extreme of these, peptides of 20–100 amino acids encoded by random fragments of *E. coli* DNA were tested for their ability to substitute for the GAL4 transcriptional-activation domain and found to be active in about 1% of cases. The only common properties in activating peptides of any origin in these experiments are a general hydrophobicity and a conspicuous excess of acidic amino acids.

Disturbing as this apparent disregard for structure may be, it is not the only recent example of a vital function with a rather loosely defined structural basis. Sequences that will direct the secretion of newly synthesized proteins can also be generated at random, with a success rate of about one in five¹⁷; and proteins can be targeted to mitochondria by 1–3% of sequences encoded by random fragments of bacterial DNA, or 5% of sequences from within a cytosolic protein¹⁸.

In the circumstances, it may be some consolation that specific sequences do seem to be required for the repression of activation by GAL4. GAL4, which controls the expression of the enzymes required for galactose metabolism, binds constitutively to DNA and is repressed in the absence of galactose by a regulatory molecule, GAL80, which binds to GAL4 with an affinity that is itself regulated by galactose or its metabolites. Ptashne reported deletion experiments by Ma showing that the 30 most carboxyl amino acids of GAL4 are essential for GAL80 binding.

Further consolation may be drawn from the reflection that biological reliability in the face of structural licence at least makes it easier to see how evolution might have occurred. Indeed, Struhl has argued¹⁹ that the combination of flexible protein domains and DNA looping may have been crucial to the evolution of eukaryotic control mechanisms.

There are two ways in which DNA looping might be important in gene regulation. It could simply allow cooperative binding between proteins at widely separated sites on the DNA and thus permit gene regula-

tion by manipulation of the concentration of a single regulatory protein, by analogy with λ repressor, or by different combinations of regulatory molecules, as might be expected in tissue-specific gene regulation in eukaryotes (see ref. 20). Alternatively or additionally, it might contribute to a specific topology of the protein-DNA complex necessary for transcriptional regulation²¹. But although DNA-protein interactions are better understood for λ repressor than for any other system, λ repressor does not in nature act at a distance. Many people are now attempting to extend the analysis of DNA looping to systems in which action at a distance does occur in life.

Looping and cooperation

The *lac* repressor of *E. coli*, for example, has two naturally separated binding sites that have recently been implicated in the cooperative repression of the *lac* promoter²², and Benno Müller-Hill (Cologne University) described experiments in which deletion of one of the two operators decreases repression²³ and in which an artificial construct containing two operator sites at varying distances on linear DNA has been used to demonstrate, both directly by electron microscopy and indirectly by increased DNase hypersensitivity of the intervening DNA, that looping is induced *in vitro* by contact between *lac* repressor tetramers²⁴. Even these experiments, however, do not so far bear directly on the mechanism of regulatory cooperation by binding at a distance *in vivo*, where the DNA is not linear; it may be relevant that Müller-Hill has now shown significantly enhanced cooperative binding in supercoiled DNA. In higher eukaryotes too, the investigation of protein-protein interactions at a distance is still limited to the search for their diagnostic signs.

There are three ways of diagnosing such interactions. The simplest to understand is electron microscopy, which can reveal loops formed in DNA by two bound proteins (or by one protein bound to two sites, which may explain looping in *lac* and in the *E. coli* *deo* operon²⁵ but has never been demonstrated). More complex is the measurement of changes in DNase hypersensitivity with distance between protein-binding sites (Fig. 1b). This works only for proteins that bind on one side of the DNA. The third test is to measure the effect of a protein bound at one site on the binding affinity of a protein bound at a second distant site. The functional consequences of cooperative binding can be inferred from changes in gene expression consequent on the addition or deletion of upstream binding sites for regulatory proteins.

By various subsets of these criteria, three demonstrations of cooperative interactions at a distance in eukaryotes

were presented at the meeting. Carl Wu (National Institutes of Health) showed that a purified heat-shock activator protein of *Drosophila* will bind with higher affinity where it has three adjacent binding sites than where there are two or one. Steven McKnight (Carnegie Institute, Baltimore) reported the isolation of a DNA-binding protein that will recognize both a promoter and an enhancer region in the long terminal repeat of the murine sarcoma virus²⁶, and showed that mutations in the promoter that affect the binding affinity of the protein also affect the affinity of the protein for the upstream enhancers, implying cooperativity. The functional relevance of these observations is established by the demonstration that transcriptional activation by the purified protein is depressed by deletion of the enhancer sequences.

On functional criteria only, Günther Schütz (Cancer Research Centre, Heidelberg) has shown that the tyrosine aminotransferase gene of mouse is controlled by the cooperative interaction, presumably through glucocorticoid receptor binding, of two glucocorticoid-response element sites 2.5 kilobases upstream of the transcriptional initiation site. Although both sites will bind purified receptor *in vitro*, deletion of the more distal site abolishes the glucocorticoid-induced transcription; glucocorticoid inducibility survives the deletion of the more proximal site but is increased in the presence of both²⁷.

A glimpse of what this nascent understanding of the molecular minutiae of gene regulation may mean in more global terms will be offered in the second part of this article, which will deal with the regulation of genes controlling the morphogenesis of the fruitfly. □

1. Ptashne, M. *Nature* **322**, 697 (1986).
2. Anderson, J.E., Ptashne, M. & Harrison, S.C. *Nature* **326**, 846 (1987).
3. Koudelka, G.B., Harrison, S.C. & Ptashne, M. *Nature* **326**, 886 (1987).
4. McLarin, J.A. *et al.* *Science* **234**, 1526 (1986).
5. Liu-Johnson, H.-N. *et al.* *Cell* **47**, 995 (1986).
6. Richmond, T. *Nature* **326**, 18 (1987).
7. Klug, A. & Travers, A. *Nature* **327**, 280 (1987).
8. Ptashne, M. *A Genetic Switch* (Cell Press & Blackwell Scientific, Cambridge, Massachusetts, 1986).
9. Schewitz, R.W., Otwinowski, Z., Joachimiak, A., Lawson, C.L. & Sigler, P.B. *Nature* **317**, 782 (1985).
10. Zhang, R.-g. *et al.* *Nature* (in the press).
11. Buc, H. *et al.* in *RNA Polymerase and the Control of Transcription* (ed. Reznikoff, W.) (Elsevier, New York, in the press).
12. Hochschild, A. & Ptashne, M. *Cell* **44**, 681 (1986).
13. Dunn, T., Hahn, S., Ogden, S. & Schleif, R. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5017 (1984).
14. Brent, R. & Ptashne, M. *Cell* **43**, 729 (1985).
15. Hope, I.A. & Struhl, K. *Cell* **46**, 885 (1986).
16. Ma, J. & Ptashne, M. *Cell* **48**, 847 (1987).
17. Kaiser, C.A. *et al.* *Science* **235**, 312 (1987).
18. Baker, A. & Schatz, G. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
19. Struhl, K. *Cell* **49**, 295 (1987).
20. Schleif, R. *Nature* **327**, 369 (1987).
21. Echols, H. *Science* **233**, 1050 (1986).
22. Mossing, M.C. & Record, M.T. *Science* **233**, 889 (1986).
23. Eisman, E. *et al.* *J. molec. Biol.* (in the press).
24. Kramer, H. *et al.* *EMBO J.* **6**, 1481 (1987).
25. Dandaneil, G., Valentin-Hansen, P., Larsen, J.E.L. & Hammer, D. *Nature* **325**, 823 (1987).
26. Johnson, P.F. *et al.* *Genes Devel.* **1**, 133 (1987).
27. Jantzen, H.-M. *et al.* *Cell* **49**, 29 (1987).

Miranda Robertson is Biology Editor of Nature.

New generation databases for molecular biology

SIR—Databases, particularly those of DNA and protein sequences, play an important part in modern molecular biology. When a new protein is sequenced, these databases allow a rapid search for homologous proteins. Molecular biology has advanced so rapidly that it may now be time to develop 'second generation' databases. Instead of storing linear sequence information, these would emphasize concepts and relationships, for example:

(1) A database of protein-DNA interactions might contain the amino-acid sequences of DNA-binding proteins and the DNA sequences of their affined binding sites. This information could be organized at different levels of detail, so that information about the precise size of the binding site, about mutations that affect binding, and even atomic details of the interaction could be added.

(2) A database that, organized with a system of 'pointers' for cross-referencing, might keep track of all known sequence homologies, along with information about the significance of each sequence match. It might also compare protein and DNA homologies.

(3) Databases of structural motifs (similar to the one started by Blundell *et al.*, *Nature* **326**, 347-352; 1987) might contain all the structural units that had been observed in more than one protein or all sequences that are homologous to proteins of known structure.

There is also a need to develop libraries of subroutines or programs that are written in different laboratories and yet can be used interchangeably. Programs using common protocols or data structures might provide the best way to handle information from a project to sequence the human genome. Agreeing on programming protocols would help to make software units compatible, just as standardization in hardware design may allow different computers to work together.

What makes a database useful? In a very general sense, the use of a database may be determined by the amount of information, the accuracy (how many mistakes were made in sequencing the protein and entering the data), and the relevance of the data to a particular problem. Sequence databases clearly satisfy these criteria — they contain a large amount of highly reliable information that is relevant to problems involving protein structure, function and evolution.

Versatility is another central issue in database design. Databases can be used in many different ways because data and analysis are clearly separated. The sequence data are permanently stored on disk or tape; many different programs can access the same database, using the

sequence information in completely different ways. Relatively little interpretation and judgement are needed as the database is developed. The more critical problems of interpretation and judgement are deferred until the sequence data are analysed with a particular program — the appropriate strategy will depend on the goals of the project, and criteria for matches are a matter of scientific judgement.

How would higher order databases differ from existing databases? Sequence data are relatively easy to organize because the natural linear structure of the information simplifies information storage. New databases, however, may be highly branched. In a database of protein-DNA interactions there might be a protein that binds at many sites; in a database of protein homologies, a protein could be related to many other proteins. Such branched structures will be inherently more difficult to organize and search than linear sequence information. (Presumably this is one of the reasons that a database of carbohydrate structures — see *Nature* **324**, 208; 1986 — has only recently been organized.)

There are additional, more fundamental problems involved with second generation databases. Establishing them will require many difficult scientific judgements. One must choose the most important relationships and concepts to include in the database, and prescient decisions will be needed if one is to develop a useful tool for scientific research. Data entry will

also be difficult. Data and analysis will not be clearly separated, and constant decisions will be required as information is entered. Data will have to be evaluated with the same care that is used when reviewing a manuscript. Finally, the versatility of these new databases must be a major concern. Will they be able to incorporate new types of information and to be used in unanticipated ways?

Is it feasible to establish higher order databases? Which will be most useful? Should they be started as collaborative ventures or should they be started in individual research laboratories and be used by other groups only after they have proven their utility? Can one really anticipate the needs for particular databases and foresee the best ways to organize them? Unlike physics, which moved to an era of 'big science' because of the costs of equipment for high energy research, it may be information that drives molecular biology into big science, and leads to a cooperative style of research. Organizing data at high levels of abstraction may be a step towards the more widespread use of artificial intelligence programming methods in molecular biology, and these higher order databases should be useful 'knowledge sources' for future programs.

CARL O. PABO

Howard Hughes Medical Institute,
Department of Molecular Biology
and Genetics,
Johns Hopkins University School of
Medicine,
Baltimore, Maryland 21205, USA

Our mistake

SIR—Bourne¹ is correct in suggesting that a problem exists with data presented in our paper². The data as presented in Figs 1 and 2 were compromised by a serious error in calculating the specific activity of GTPγS. The correct values for GTPγS binding should be 1,000-fold lower than shown in our paper; that is, the y axis of the left-hand side of Fig. 1 should be fmol, not pmol, and the x axis of Fig. 2 should be pmol mg⁻¹, not nmol mg⁻¹.

We apologize if the data presented in our paper misled any investigators and accept full responsibility for this error. We are confident that the observations reported are real and that they demonstrate a role for a GTP binding protein in IL-2 mediated signal transduction.

STUART W. EVANS
SUZANNE K. BECKNER
WILLIAM L. FARRAR

National Cancer Institute,
Division of Cancer Treatment,
Frederick Cancer Research Facility,
Frederick, Maryland 21701, USA

1. Bourne, H.R. *Nature* **326**, 833-834 (1987).

2. Evans, S.W., Beckner, S.K. & Farrar, W.L. *Nature* **325**, 166-168 (1987).

Chromaffin cell synapsin?

SIR—In two recent issues of *Nature* there were reports with accompanying News and Views items on the role non-erythroid brain spectrin (fodrin) and associated cytoskeletal proteins play in the mechanisms of chromaffin cell exocytosis in the adrenal medulla^{1,2} and neurotransmitter release in the brain^{3,4}. Data discussed in these articles suggest the possibility that the two systems, both of which are derived from the neural crest, may share a common mechanism.

In the case of synaptic transmission, it was suggested⁴ that neurotransmitter-containing synaptic vesicles are restrained in a fodrin/actin network in the presynaptic terminal until nerve depolarization elevates free cytosolic calcium. This activates a calcium/calmodulin-dependent protein kinase that phosphorylates synapsin I, a synaptic vesicle-associated protein that shares a number of properties with erythrocyte protein 4.1 (refs 5,6) and protein 4.9 (refs 3,7). On phosphorylation, synapsin I, which can enhance spectrin/actin interactions only in its dephosphorylated form⁸, would lose its ability to mediate that interaction thus leading to

dissolution of the network, freeing the vesicle for exocytosis at the presynaptic membrane.

For the process of chromaffin granule secretion, it was suggested² that the secreting granules are crosslinked within a fodrin/actin cytoskeleton in the resting state and are released following network disassembly that, as in the synapse, is secondary to calcium influx, activation of calcium/calmodulin-dependent protein kinase, and the phosphorylation of an as yet unidentified granule membrane or cytosolic component⁹.

Might this component be a synapsin-like molecule? In a study directed at determining which chromaffin granule membrane proteins become phosphorylated when stimulated in the presence of [γ -³²P]ATP¹⁰, polypeptides of *M*_r 60,000 (tyrosine hydroxylase) were described. It is of interest to note, however, that careful examination of their autoradiograph reveals additional labelling of polypeptides of about *M*_r 73,000, remarkably similar to those found for synapsin Ia and Ib on discontinuous SDS-PAGE. Furthermore, in one of the early synapsin

studies¹¹, the molecule (then known as Protein I) was shown to be abundant in the adrenal medulla. Determining whether synapsin is confined to nerve terminals in this tissue or whether it is also present in chromaffin cells may allow what has been learned about neurotransmitter release in the brain to be applied to the molecular mechanism of chromaffin cell exocytosis.

DON L. SIEGEL

Department of Anatomy,
University of Pennsylvania
School of Medicine,
Philadelphia, Pennsylvania 19104, USA

1. Perrin, D., Langely, D.K. & Aunis, D. *Nature* **326**, 498-501 (1987).
2. Burgoyne, R.D. & Cheek, T.R. *Nature* **326**, 448 (1987).
3. Bahler, M. & Greengard, P. *Nature* **326**, 704-707 (1987).
4. Baines, A.J. *Nature* **326**, 646 (1987).
5. Baines, A.J. & Bennett, V. *Nature* **315**, 410-413 (1985).
6. Goodman, S.R. & Zagon, I.S. *Am. J. Physiol.* **250**, 347-360 (1986).
7. Siegel, D.L. & Branton, D. *J. Cell Biol.* **100**, 775-785 (1985).
8. Krebs, K.E., Zagon, I.S. & Goodman, S.R. *J. Cell Biol.* **103**, 2829a (1986).
9. Trifaro, J.-M., Bader, M.-F. & Doucet, J.-P. *Can. J. Biochem. Cell Biol.* **63**, 661-679 (1985).
10. Burgoyne, R.D. & Geisow, M.J. *FEBS Lett.* **131**, 127-131 (1981).
11. DeCamilli, P., Ueda, T., Bloom, F.E., Battenberg, E. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5977-5981 (1979).

Impact cratering and flood basalt volcanism

SIR—Two contrasting causal mechanisms for the mass extinctions at the Cretaceous/Tertiary (K/T) boundary at 66 Myr BP continue to be the subjects of intense debate: (1) a comet or asteroid impact (or several impacts during a comet shower)¹, and (2) increased volcanic activity, as summarized recently by Officer *et al.*². Evidence for a large-body impact is strong — a worldwide ejecta layer containing anomalous iridium, with microspherules and shocked minerals, coincides closely in time with the mass extinctions on land and in the sea³. There are also several candidate K/T craters on the continents, although none so large (100 to 200 km in diameter) as suggested by the composition of the impact layer. On the other hand, recent study shows that the great Deccan flood basalts in India were erupted in a period of only 1–3 Myr, also coinciding with the K/T boundary⁴, and with possible widespread environmental effects⁵.

The seeming coincidence raises the question: could flood basalt eruptions and large-body impacts be causally related? Impacts might lead to basaltic volcanism directly by penetrating and deeply fracturing the Earth's crust, allowing magma to reach the surface, or by causing pressure-release melting in the upper mantle by the excavation and removal of crustal rocks, and thus triggering a rising plume of magma⁶. Very large impacts might also indirectly trigger volcanism through seismic disturbances at pre-existing volcanically active zones, but this is very uncertain. The major question is: how deeply do

large-body impacts excavate the Earth's surface? Or more generally: what is the depth/diameter ratio for large impact craters? Two studies reported at the recent 18th Lunar and Planetary Science Conference in Houston have now provided quantitative evidence for an answer^{6,7}.

Previous work showed that small impacts form simple, bowl-shaped craters⁸ with a true depth (*d*) to diameter (*D*) ratio, *d/D* = 0.2 to 0.3 (taking into account partial fill by impact breccia⁸. At diameters above about 1–2 km in sedimentary targets and 3–4 km in crystalline targets, however, terrestrial impact craters have a complex form, with a central core of uplifted, shocked rocks surrounded by one or more concentric depressions (terraces formed by faulting and slumping of the outer portion of the crater). These features apparently represent collapse of the initial crater and isostatic rebound of the underlying crust and mantle as they adjust to the sudden removal of the surface rocks⁹. Most studies agree that the final crater form for large impact craters should therefore be quite shallow (*d/D* = 0.1 to 0.006 in some cases) and this agrees with observations for large craters on the Moon and other terrestrial planets¹⁰.

But what about the depth of the initial transient crater produced during large impacts? Although it has been claimed that large craters might be initially quite shallow (see, for example, ref. 10), with only minimal penetration on impact, two recent studies suggest that the depth of excavation of large terrestrial craters is great enough to have considerable geological effects on the lower crust and upper mantle. In the most recent detailed

modelling of impact penetration and later crater readjustment, O'Keefe and Ahrens⁶ show that for high accelerations and large scales, a deep transient bowl-shaped crater is always produced. They conclude that the initial crater morphology is given by $d_e/r = 0.002r^{0.2}$ (in c.g.s. units, where d_e is the depth of excavation, r is the crater radius, for the Earth's gravity and an impact velocity of 12 km s⁻¹). This agrees in general with earlier work by Orphal¹¹ and Melosh⁹, and indicates that large impact craters should have had deep initial penetration cavities.

If the above relationship holds true, then the transient depth of excavation of a 100–200 km diameter crater on the earth (as proposed for a 10-km diameter impactor at the K/T boundary) would be in the range of about 20–40 km; deep enough to penetrate easily through the 5-km-thick ocean crust, and even deep enough to excavate much or all of the thicker continental crust. This supports the evidence presented at the same conference by Hildebrand and Boynton⁷, which shows that the K/T ejecta layer has a substantial mantle component. Rare-earth element abundance patterns in the boundary layer can be explained by a mixture of ocean crust, continental sediments and oceanic mantle; the fraction of mantle material indicates a minimum depth of cratering of about 40 km. These results suggest that the crater produced by the proposed K/T impactor could have been deep enough to generate directly volcanism in an oceanic impact, or create deep fractures in the lower continental crust. Perhaps the target area(s) for the largest K/T impactor(s) lies beneath the Deccan Traps, the Iceland hotspot (the source of the Brito-Arctic basalt eruptions), or some other hotspot or oceanic plateau¹². The extinctions at the K/T boundary might therefore be related both to impact and volcanic phenomena.

MICHAEL R. RAMPINO

Department of Applied Science,
New York University,
New York,
New York 10003, USA

and
NASA, Goddard Institute for Space
Studies, New York,
New York 10025, USA

1. Alvarez, W. *Eos* **67**, 649–658 (1987).
2. Officer, C.B., Hallam, A., Drake, C.L. & Devine, J.D. *Nature* **326**, 143–149 (1987).
3. Courtillot, V. *et al.* *Earth planet. Sci. Lett.* **80**, 361–374 (1986).
4. Stothers, R.B., Wolff, J.A., Self, S. & Rampino, M.R. *Geophys. Res. Lett.* **13**, 725–728 (1986).
5. Green, D.H. *Earth planet. Sci. Lett.* **15**, 263–270 (1972).
6. O'Keefe, J.D. & Ahrens, T.J. *Abstr. Eighteenth Lunar Planetary Sci. Conf.* 744–745 (1987).
7. Hildebrand, A.R. & Boynton, W.V. *Abstr. Eighteenth Lunar Planetary Sci. Conf.* 427–428 (1987).
8. Grieve, R.A.F. *Spec. Pap. geol. Soc. Am.* **190**, 25–37 (1982).
9. Melosh, H.J. *J. geophys. Res.* **87**, 371–380 (1982).
10. Pike, R.J. *Icarus* **43**, 1–19 (1980).
11. Orphal, D.L. *Abstr. Tenth Lunar Planetary Sci Conf.* 949–951 (1979).
12. Rodgers, G.C. *Nature* **299**, 341–342 (1982).

Only connect!

Richard Morris & David Willshaw

Parallel Distributed Processing: Explorations in the Microstructures of Cognition.

By David E. Rumelhart, James L. McClelland and the PDP Research Group.

MIT Press: 1986. Vol. 1 Foundations pp.547; Vol.2 Psychological and Biological Models pp.611. \$27.50, £24.75 each volume.

THIRTY years ago, Karl Lashley concluded his unsuccessful efforts to discover where memories are localized in the brain with the famous cry of despair that "learning is just not possible". His experiments showed that the performance of rats in learning complex mazes was deleteriously affected by lesions made anywhere on the cerebral cortex, and that the effects were in proportion to the lesions' size. These two findings were enshrined in the principles of "equipotentiality" and "mass-action". Later work has revealed the exquisite localization of function to specific areas of cortex, but the sentinel of mass-action has stood, albeit dented and bruised, over the years. It is now finding expression in an exciting new approach to the understanding of cognition.

Quite independently, many people have recently become attracted to the idea that cognitive processes reflect the simultaneous activation of a large number of simple processing units. The slow and probably imprecise functioning of individual nerve cells seems to demand that they work collectively in parallel. Such systems would continue to function with incomplete or inaccurate inputs, so having the properties of content-addressability and generalization, and thus be compatible with mass action. These two volumes describe recent research by two cognitive psychologists and their colleagues into models of parallel distributed processing (PDP). Their work is relevant to cognitive psychology, neurobiology and, although not discussed explicitly, parallel computing.

A PDP model consists of a network of interacting simple units. Its overall behaviour is determined by the levels of activation of the units and the weights of the individual connections between them. For the network to act as an information-processing system, the weights must be set so that a given stimulus applied to the input units produces the correct output. A key question for researchers has been to find procedures for the correct setting of the weights for a given task.

A learning machine of this type called the "perceptron" was exhaustively analysed by Minsky and Papert in the 1960s (*Perceptrons*; MIT Press, 1969). They showed that the single-layer perceptron could do rather little, but they recognized that multilayered perceptrons, containing

"hidden units" at levels intermediate between input and output, were potentially very powerful. Only recently have learning algorithms been discovered for systems of this type, and Vol. 1 contains formal analyses of them, including two radically new ones which are currently the subjects of great interest.

In both of the novel learning algorithms, the procedure for changing the weights has the effect of minimizing an energy function, which corresponds to the total error in the machine's output and which has a value for each possible configuration of weights. In the Boltzmann machine, inputs are presented in turn and a local learning rule modifies each weight according to how often a unit is active compared with how often it should be active. Such a method, which involves gradually changing some measure of error, frequently finds local minima of energy corresponding to sub-optimal solutions. To circumvent this problem, the method of simulated annealing is used, analogous to the way that a bouncing ball reaches the bottom of a hill while a rolling ball may not. In the second algorithm, the error at the output stage is propagated back to the hidden units. This is done because the desired states of the hidden units are not known. The back-propagated error is then used to modify the weights onto the hidden units so as to minimize the total error in the network. Many of the computations which frustrated Minsky and Papert are performed with surprising ease using these new algorithms.

PDP models would be of little more than academic interest were it not for their implications for both psychological and neurobiological models of cognition. These issues are addressed in the second volume, which opens with that old warhorse of cognitive science — "schema" — seen in the PDP framework as a "higher level" concept, emerging out of the interactions of activated units. Various psychological models are presented: of reading, of the perception of speech, of past-tense learning and of sentence processing. Discussion of a PDP model for the learning of past-tenses brings out a central theme — that a cognitive system based on PDP can behave as if it has learnt an explicit rule "under the guidance of some explicit overseer", even though the learning has

actually involved no more than changing connection strengths between units on the basis of locally available information.

These psychological models are based on relatively simple PDP models first discussed in the 1970s by Anderson, Kohonen and others. Their performance seems to depend heavily on the method of encoding the raw data, which is done by hand, rather than using the new algorithms described in Vol. 1. For example, in the distributed model of memory (Chapter 17), the successful extraction of prototypes (*dog*, for example) seems to be accomplished as much through the particular method of encoding exemplars as in the behaviour of the memory system itself. For this reason, the psychological section may disappoint some readers, and we await the promised application of the new algorithms. More attention could have been given to whether the logical relationships between the items under consideration are consistent with the constraints imposed by the computational machinery. For example, with respect to generalization to a novel stimulus, that furnished by a simple linear model will be different to that by a linear model with thresholds — a distinction made incorrectly throughout this work (linear independence versus linear separability).

Constraints arising from what is biologically reasonable are not given much attention — so much so that Crick and Asanuma, in a chapter surveying relevant aspects of the anatomy and physiology of the cerebral cortex, complain that "certain devices loved by theorists" have not yet been identified. To the extent that units are to be identified with neurons (on which the authors hedge their bets), this problem besets several of the formal models. Sejnowski adopts a broader view of biological constraints, and makes several questionable assertions, among them that memory and information processing occur within the same cortical circuitry. Unfortunately, there is still no clear evidence on this point. Zipser outlines a simple PDP model of place recognition, inspired by O'Keefe's discovery of "place-cells" in the hippocampus. It is pleasing to see a model based on what the brain can do rather than on what it might do. McClelland and Rumelhart conclude with a model of amnesia which is curious because it is built around the temporal gradient of retrograde amnesia — neither a defining nor an obligatory feature of the syndrome. The model predicts that amnesics forget faster than those with normal memory, an idea from which neuropsychologists, if they ever did believe it, are hastily retreating.

A puzzling omission is the lack of reference to the important phenomenon of long-term potentiation (LTP) in the

hippocampus, a physiologically induced increase in synaptic efficacy. LTP occurs under conditions which closely resemble the Hebbian rule for synaptic modification ("no theorists continue to use it . . .", p. 37, Vol. 1) and is known, from behavioural studies, to be of functional relevance. Omission of any mention of LTP seems strange given the emphasis placed upon the idea that knowledge resides in the *connections* between units; and this Hebbian blind spot mirrors the confusion shown in Vol. 1 over the relation between an error-correcting learning rule (which requires a teacher) and a Hebbian rule (which does not). Here, surely, is a candidate mechanism for altering weights between units; and that is what PDP modelling is all about.

This is a stimulating treatise which will remain an important source of material for some time. However, there are errors

and omissions which display the authors' lack of familiarity with related work; the chapters are interlinked like a cat's cradle, making them difficult to read; and the reader may be put off momentarily by the repetitive and over-ambitious claims of some authors. Some critics of PDP are concerned about whether the approach is at an appropriate level of explanation for understanding cognitive processes. The authors tell us that "the real proof is in the pudding", others see a revolution in the making. Our view is that real progress will require further development of PDP algorithms, investigations of their computational power and consideration of their experimental implications by both neurobiologists and psychologists. □

Richard Morris is in the Department of Pharmacology and David Willshaw the Department of Zoology, University of Edinburgh, Edinburgh, UK.

Hazardous advice

David E.F. Harrison

The Environmental Risks from Biotechnology. By M. Chiara Mantegazzini. Frances Pinter, London/Columbia University Press, New York: 1986. Pp.164. £17.50, \$30.

Biotechnology Risk Assessment: Issues and Methods for Environmental Introductions. Edited by Joseph Fiksel and Vincent T. Covello. Pergamon: 1986. Pp.174. \$27.50, £19.50.

WE HAVE come a long way since 1974, when *Nature* and *Science* published a call by Professor Paul Berg and colleagues for a moratorium on genetic engineering. Our understanding of the risks presented by biotechnology is now much greater, but, as scientists, we cannot afford to be complacent. The public has a right to question and investigate the long-term effects of scientific and technological activities, and the onus is on scientists and technologists to prove beyond reasonable doubt that any real risks can be managed. The logical and desirable situation is to be able to base regulations and guidelines on a system of accurate risk assessment. Until now, the tools for assessment of the risk from recombinant organisms in the environment have not been readily available to researchers, and any volume which goes a way to providing them should be most welcome.

Superficially, these two books are very similar — not only do they address the same topic but both were written as reports for government agencies. Dr Mantegazzini's was prepared for the Commission of the European Communities Director-General Environment, Consumer Protection and Nuclear Safety, while

Fiksel and Covello's was for the US Office of Science and Technology Policy Executive. However, the similarity does not go much further than that. Presumably each report fulfilled the requirements of the agency that commissioned it, and from them one is tempted to draw comparisons between the European and American bureaucratic styles. Here, however, I will confine myself to reviewing the books' contents.

The American report is a collection of eight papers by twelve authors, clearly aimed at a scientifically sophisticated readership, while the European volume was prepared by one person and, from its style, one would assume is intended for less well-informed readers. There is also a subtle difference in the stated scopes of the two books: the American work limits itself to a consideration of the methodology of risk assessment while the European report sets out "to evaluate risk to the environment of products and processes in the field of biotechnology". It does, in fact, go even further and tries to define biotechnology, identify the possible fields of application and so on. This was a most ambitious goal in such a short book, almost half of the 164 pages of which are taken up cataloguing and discussing European Community regulations on the subject.

Given the restricted length Mantegazzini could only cover the general issues superficially, and the only part of her book likely to be useful to biotechnologists is that reviewing European regulations. However, the book could well be worth purchasing for this alone. There is a short list of conclusions and recommendations — all good but general stuff, for example "general guidelines can be worked out . . ." and "Deliberate release of micro-organisms requires a previous case-by-case assessment of potential hazards"

(what if there has been no previous case?) — but these obviously leave plenty of scope for further reports and for employment of committees.

At the end of the book an "Ad-Hoc Directive" is proposed which again is somewhat superficial and limited in scope, and gives the impression of having been reduced to the uncontroversial minimum. Among the list of recommended procedures is that information should be submitted on "the foreseeable risks immediate or delayed, for man and environment". But that, of course, begs the question of how such risks can be determined.

This is where Fiksel and Covello's book steps in. It is a collection of thorough and impressive analyses by experts, and covers almost every aspect of evaluation of risk to the environment from biotechnology. Topics covered in depth include modelling of risks; assessment of risks from bacteria, fungi and viruses to human beings and the environment; how micro-organisms are spread; and how risks can be monitored. A great deal of practical information is included, and the book is a must for those involved in any way in the release of recombinant organisms into the environment. It will also be of great value to anyone with an interest in the assessment of safety of microbial systems in general.

The American report is clearly one on which action can, and is intended to be, taken. As with the European report, there is a list of items to be considered for submissions for the release of recombinant micro-organisms into the environment. This list, however, is twice as long and is very detailed and precise in its list of required information. It will no doubt form the basis of requirements for all future submissions in the United States and elsewhere.

The debate on release of recombinant organisms into the environment still has some way to run. Both of these volumes have a contribution to make in putting it on a more informed basis. Anyone with an active interest in the subject will find the Fiksel and Covello work invaluable, while for the lay person wanting a broad overview, Mantegazzini's book would be a reasonable starting point. □

David E.F. Harrison is Managing Director of Biotechnology Australia Pty Ltd, 28 Barcoo Street, Roseville, New South Wales 2069, Australia.

New in paperback

A classic of the literature of physics, N.F. Mott and H.S.W. Massey's *The Theory of Atomic Collisions*, is available in paperback for the first time. The book first appeared in 1933 (it cost 17s 6d and was reviewed in *Nature* by H.A. Bethe). A second edition came out in 1949 and the third in 1965. The paperback is a reprint of the third edition and is in two volumes, price £12.50 and £15 (provisionally \$19.95 and \$27.50 in the United States). Publisher is Oxford University Press.

The messages that matter

Eric Ashby

The Bird of Time: The Science and Politics of Nature Conservation. By N.W. Moore. Cambridge University Press: 1987. Pp.290. Hbk £27.50, \$49.50; pbk £9.95, \$15.95.

NORMAN MOORE's first encounter with DDT was in a German prisoner of war camp in 1945. "Before returning to polite society", he wrote, "we had to be deloused. In an environment in which typhus was active, I was delighted to be liberally sprinkled with DDT . . .". Fifteen years later he was appointed to create and run a Toxic Chemicals and Wildlife Section, where research on the effects of pesticides and herbicides on the environment could be done for the British Nature Conservancy.

For the rest of his career Moore studied the unwelcome side-effects of organochlorine insecticides on birds and mammals and benevolent insects. He patiently reasoned with government bureaucrats and farmers' lobbies to persuade them that the spread of this new kind of agricultural technology was becoming a threat to the balance of nature. And he wrote popular articles and addressed meetings in the hope that people would come to regard conservation of habitats and the protection of wildlife as essential concerns in a civilized community. In *The Bird of Time* he pauses to look back on a lifetime spent caring for the countryside.

The subtitle is *The Science and Politics of Nature Conservation*. But this is not (in my view) what the book is about. Anyone who wants to read a documented story about the politics of pesticides in Britain has to turn to John Sheail's masterly *Pesticides and Nature Conservation* — it is incomparably better for this purpose. And anyone who wants to read about the scientific aspects should first consult Kenneth Mellanby's *Pesticides and Pollution*. Dr Moore's book is rather the personal record of a scientist whose work is the foundation upon which our still rudimentary policies for conservation are built, and it is for this that it is worth reading.

Moore reminisces about his work as regional officer for the Nature Conservancy in south-west England. It is not difficult to see how his career became set on conservation. Figure 13 of the book carries four ominous maps, showing the shrinkage of Dorset heathlands from 1896 to 1960, due to the inroads of farmers and builders; and figure 16 shows how the hedges surrounding fields in a part of East

Anglia all but disappeared between 1946 and 1965. Heaths and hedges carry a rich community of living things. If the hedges are uprooted and the heaths used for crops or conifer plantations or urban development, the living communities vanish. So the choice of nature reserves and the preservation of what were called "sites of special scientific interest" (SSSIs) were concerns to which Moore willingly devoted his skill.

Nature reserves and SSSIs take land which farmers could use for profit. So there have to be compromises; farmers have to be won over to the beliefs of conservationists. Moore was well qualified to negotiate with farmers, for he understands farming himself and he rejects the

governments was that birds *were* at risk, though his evidence was weakened by all sorts of reservations and his case could not be conclusive. The simple concept, that all birds at the top of the food chain accumulate organochlorine residues, is not correct. Yet once such reservations are mentioned to officials or to the public, the issue becomes confused and governments are virtually invited to wait until more research is done. So Moore, like the rest of us, had to learn, as he puts it, "the power of simple concepts irrespective of their accuracy", though in his lectures he was at great pains to emphasize the inconsistencies in the evidence and the difficulty, in this kind of biology, of designing experiments to prove a point.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Deadly diet? A peregrine falcon, the top of a food chain, with its prey.

common assumption that farmers are averse to conservation. One of his valuable contributions to the diplomacy of conservation was as a founder member of the Farming and Wildlife Advisory Group (FWAG) which enlists the co-operation of farmers who want to manage their land in ways that preserve natural habitats. His experience with FWAG was invaluable when, later on, he had to convince farmers that persistent pesticides, which had brought them great benefits, were putting at risk the very survival of peregrine falcons, *and that this mattered*.

These last four words were the hardest part of the message to get across. Moore's account of the difficulty of doing this is the most interesting part of the book, for part of the difficulty lay in his own scientific integrity. It has been said of him that "his style and manner resembled that of the university don". In the book he carefully distinguishes between scientific proof and circumstantial evidence; his dilemma was that time was running out for the birds at risk from pesticides, and his message to

The Bird of Time is written in an informal relaxed style, like leisurely after-dinner talk, moving from one episode in Moore's career to another, and coming back, after a few pages, to the same basic beliefs. Look at a hedgerow in the time-scale of evolution, not of a man's lifetime; to preserve habitats is paramount — if this is done well, communities may survive; it is essential to monitor the consequences of any new substance released into the environment; anything which endangers wildlife of "no commercial value" may in the long run endanger ourselves. It is the *Apologia* of a biologist who has been (as he puts it) "prepared to advise on the strength of the available evidence" even if it is circumstantial and vulnerable to criticism, and whose reasonableness has in the long run proved to be more lastingly effective than the strident protest of activists. □

Lord Ashby is a Fellow of Clare College, Cambridge CB2 1TL, UK. He was Chairman of the Royal Commission on Environmental Pollution from 1970 to 1973.

Clash of symbols

Peter Kemp

Einstein's Monsters. By Martin Amis. Jonathan Cape, London/Crown, New York: 1987. Pp. 127. £5.95, \$12.95.

BROODING OVER nuclear weapons has made Martin Amis's imagination mutate, he claims. Breeding a family has played its part too. As his introduction to the five fables collected in *Einstein's Monsters* explains, perusal of Jonathan Schell's grim forecast of nuclear holocaust, *The Fate of the Earth*, alerted him to looming danger, while the birth of his two children — awakening protective instincts in a writer who earlier turned out ostentatiously callous novels with titles such as *Dead Babies* — only increased his anxiety.

Belatedly becoming aware of the nuclear menace in 1984 at the age of 35, Amis now proclaims it with near-adolescent raucousness. Previously keen on presenting himself as a knowing habitué of the lurid, high-risk zones of London or New York, he uses this book to flash his credentials for admission to the scary sphere of nuclear nightmare. His opening polemic — couched in his customary style of finger-clicking, hard-boiled swagger enlivened by some graphic phrasing — retails familiar facts about the horrendous potentialities of nuclear armaments. Some undocumented claims, as that stockpiled weapons emit radiation causing cancers, are thrown in too.

The most malign emanations from the missile silos though, are, Amis fancies, contemporary crimes of violence and sexual outrage. Nuclear armaments, "Einstein's monsters", have, he claims, made monsters of us all. Their very presence has warped humanity, generating pornography, muggings, child rape and infanticide.

This unusual objection to the megamissiles features prominently in two tales here which struggle to link twentieth-century degeneration with nuclear arsenals. In one story, telling how a tough guy abstains from retaliation on two thugs who have randomly slaughtered his mother, daughter and granddaughter, Amis rather desperately strives for nuclear reasonance by endowing his pugnacious hero with "a neutronium fist", and sprinkling the prose with words such as "fallout" and "deterrent". Similarly strained is the symbolic background to the next piece — a forest buzzing with the "radiation" of insects, a lake oddly likened to a "warhead" — against which the schizophrenia and suicide of a nuclear physicist's son are portrayed.

The last three stories are more surreal and envisage the Earth after nuclear

catastrophe. "The Little Puppy That Could", a version of the Perseus-Andromeda myth, does this most extensively, evoking a garish, rotting world of hideous mutants. Many human beings are now flipped or furred. Rule is by sexually voracious matriarchal viragos with names like Keithette or Royene. Terrorizing their settlement near a still-flaming nuclear crater is a giant dog with "carcinogenic teeth", "fungoid" hair, "malarial eyes", and poisonous scarlet saliva which dissolves flesh and bone. As Amis raptly itemizes the rippings and

ravagings of this grisly beast festering with anthrax, yaws, serpigio, glanders and mange — or elsewhere pores over panoramas of a mouldering, defiled planet — he reveals how little he has really changed. Like some monstrous extension of the urban cesspits that earlier engaged his interest, the prospect of terminal global pollution finally just stimulates him into another bout of his favourite fictional game: playing with putrescence. □

Peter Kemp, 61 Princes Avenue, Finchley, London N3 2DA, UK, is an Associate Lecturer in English at Middlesex Polytechnic.

One in the eye

Michael Nelson

The Potato in the Human Diet. By Jennifer A. Woolfe. Cambridge University Press: 1987. Pp. 231. £17.50, \$32.50.

THE humble potato may lack the sophistication of the asparagus or the brashness of the leek, but it is firmly loved throughout the world and has an important role in human nutrition. Jennifer Woolfe has produced a remarkably readable commentary which confirms and highlights the potato's many virtues.

Following a description of the tuber components (all is not as uniform in a potato as it seems), Ms Woolfe devotes the second chapter to an extensive consideration of the nutritional value of the tuber in the forms in which it is commonly eaten. Numerous, well-constructed tables facilitate comparison of the nutrient content of the potato with staple roots and cereals and other vegetables. Discussion of the contribution of the potato to nutrient intake is uneven in its coverage, however, and might have been better as a single, more comprehensive chapter.

Chapter 3 describes potato protein quality and reviews the contribution that potato protein can make to the maintenance of nitrogen balance in both adults and children, especially in developing countries where potato is an important staple; here the author exhorts planners to consider the implications of switching to high-yield varieties with reduced protein levels. It is not clear at whom the final section on extraction of protein from potato processing waste is directed, as the process is too expensive to be of value in developing countries and too fraught with problems to be of much commercial interest elsewhere.

One-third of the book (Chapter 4) is devoted to the effects of storage, cooking and processing on the nutritive value of potatoes. This is an excellent, comprehensive review of the subject. The section dealing with domestic preparation and cooking losses will be especially valuable

for those involved in nutrition education, particularly in the Third World, while that on commercial processing includes helpful summaries of large-scale processes with references for further details. The account of ways to improve yield and storage in Third World countries does not always include consideration of appropriate technology (for example irradiation to delay sprouting), and information on the increase in nutritional value of certain products with addition of vitamin C (instant mashed potato) or polyunsaturated fatty acids (crisps) would have been useful.

The build-up of toxins after harvesting, an important topic, is dealt with in Chapter 5, which covers the whole range of anti-nutritional factors and their potentially damaging effect on human health in areas where potato is a staple food. The book concludes with an interesting chapter on patterns of potato consumption in the tropics, with hints on how consumption can be estimated in different sections of the population.

The book is pleasingly illustrated, well laid out, thoroughly referenced and well indexed. Occasionally it reads like sales propaganda from the International Potato Center (the funding body), though this is not too obtrusive. One drawback is that no history is included, nor is there a section on trends in world potato production and consumption by region or country. Altogether, although it is not quite in the league of the FAO/WHO monographs on wheat and legumes, the book is a valuable contribution to the literature on food commodities. Nutritionists, dietitians, food planners and others with an interest in food production and supply will find it well worth reading. □

Michael Nelson is a Lecturer in the Department of Food and Nutritional Sciences, King's College London (KQC), Campden Hill Road, London W8 7AH, UK.

• More than compensating for the absence of history in Jennifer Woolfe's book is *The History and Social Influence of the Potato* by Redcliffe Salaman. The book was first published by Cambridge University Press in 1949 and was re-issued in 1985, price pbk £13.50, \$17.95.

Prospects for the control of AIDS by immunizing seropositive individuals

Jonas Salk

The long incubation period between infection and the development of clinical AIDS may be due to an immune response to the initial infection which persists with health and wanes with disease. If this response can be boosted, it may be possible to reduce the viral burden, prevent the development of disease and reduce contagiousness.

THE purpose of this communication is to suggest a way of thinking about the phenomenology of the acquired immunodeficiency syndrome (AIDS^{1,2}) and its causative virus (the human immunodeficiency virus, HIV)³⁻⁵, and to propose an approach to their control by immunizing HIV seropositive individuals.

With most viruses the immune response eliminates the infectious agent from the host. But some viruses remain latent, for example the herpes group, which persist in nerve cells in a latent form even after an immune response is expressed and, from time to time, become activated. When this occurs newly formed viruses or viral proteins are released at nerve endings, causing a lesion; the cause of this lesion is immunopathic (host determined immunopathology) as well as cytopathic (virus determined cytopathology)^{6,7}. A similar pattern is seen with hepatitis B virus which persists in liver cells in an active or latent form; however, in this example, immunopathicity (chronic inflammation) occurs without cytopathicity (liver cell destruction), as seen in patients with chronic hepatitis^{8,9}. A comparison of different patterns of cytopathicity, latency and immunopathicity for the aetiological agents of six viral diseases is shown in Table 1.

Retroviruses as a group also cause disease by more than one mechanism. Some induce chronic inflammation and others immunosuppression, presumably through an immunological reaction to viral proteins present on cell surfaces, as well as by virus-induced cytopathology. The chronic inflammatory arthritis in goats, caused by caprine arthritis-encephalitis virus (CAEV, a lentivirus), is the result of a delayed-type immunopathological response to CAEV viral proteins^{10,11}. Immunosuppression in AIDS, caused by HIV (also a lentivirus), may be attributable to an immunopathic response to viral proteins on the surface of lymphocytes that carry the antigen CD4, as well as to a cytopathic effect of virus expression in these cells¹².

In HIV infection, as with other retrovirus infections, genetic information for replication of virus or viral elements becomes part of the genome of the in-

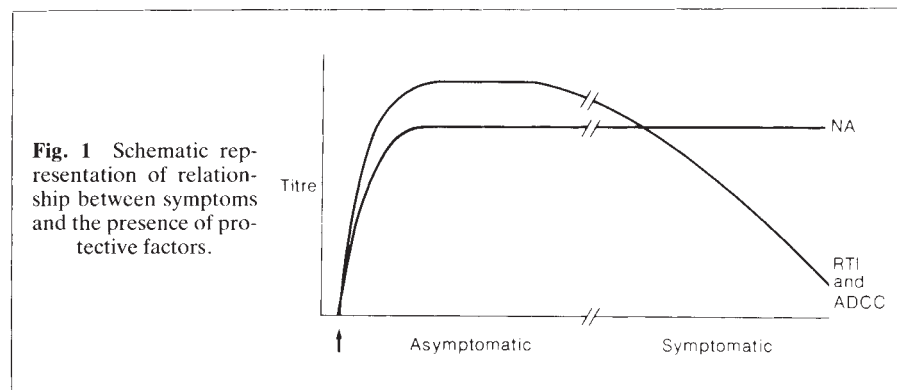


Fig. 1 Schematic representation of relationship between symptoms and the presence of protective factors.

fecting cell¹³. The observation that the time interval between HIV infection and development of AIDS can be several years suggests that this virus establishes a latent state characterized by little or no virus replication and without clinical signs of disease. Conversion of a latent asymptomatic to a productive HIV infection with clinical sequelae may be due to the stimulation of CD4⁺ cells harbouring HIV proviruses by a variety of agents, including infection with cytomegalovirus or Epstein-Barr virus¹⁴⁻¹⁶. Once activated, the pathological effect appears to be due not only to a cytopathic effect of virus replication but to an immunopathic response to cell-produced or cell-associated viral proteins^{12,16}, analogous to a 'host-versus-graft' type reaction.

Another view is that immunoprotective factors induced by the initial infection persist in those who remain healthy and are lost or are overwhelmed by superinfection in those who develop AIDS^{17,18}. If so, what may these factors be and can some form of immunological intervention be devised that would, if applied early enough, prevent the development of AIDS in serologically positive individuals?

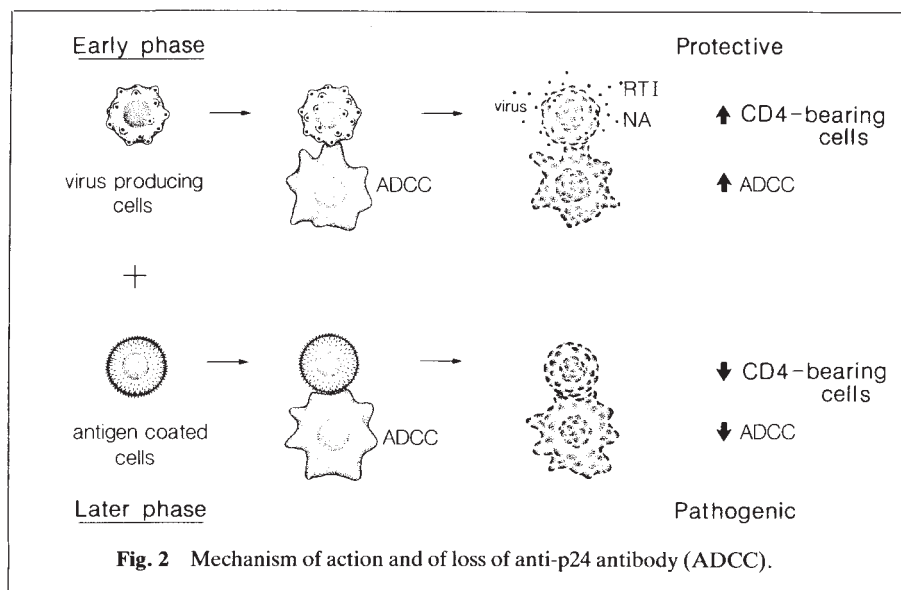
Hypothesis

The clinical syndromes (AIDS-related complex (ARC) and AIDS) associated with HIV infection are attributed by some principally to the cytopathic effect of HIV on lymphocytes that carry the CD4 antigen. An alternative view is that immunopathic factors are principally responsible for the pathogenesis of ARC and AIDS and that loss of CD4⁺ cells is a conse-

quence of host immune destruction of CD4⁺ cells as well as of viral-induced cytopathicity^{12,16}. An understanding of the mechanism whereby disease fails to develop in a healthy human carrier could provide a clue as to the form of immunological intervention that might prolong the healthy state. Such a clue may be found in a comparison of the immunological profiles of serologically positive healthy individuals with those of patients with ARC and AIDS^{16,18,19}.

As shown schematically in Fig. 1, antibody to gp41¹⁸ and virus-neutralizing antibody (NA)²⁰ which are present in the asymptomatic phase of HIV infection also persist in the symptomatic phase. But the level of anti-p24 antibody (which correlates with the presence of antibody-dependent cell cytotoxicity (ADCC)¹⁶) is present in the asymptomatic phase and declines in the symptomatic phase^{16,18,19}. Similarly, HIV seropositive sera contain an antibody that inhibits the function of reverse transcriptase (RTI); this is also present in the asymptomatic phase and declines in the symptomatic phase (ref. 21 and D.T. Imagawa, personal communication). The latter has also observed that in individuals in whom RTI antibody is present attempts at virus isolation are less frequently positive than in those in whom it is absent.

The hypothesis proposed is that, in addition to cytopathicity attributable to virus spread in the host, a further cause of the progressive pathology in patients with AIDS is attributable to the production of immunoreactive HIV proteins²². These proteins are produced in or become



attached to CD4⁺ lymphocytes, marking these cells as foreign and as targets for destruction by ADCC and other immunological cytotoxic mechanisms. When the rate of loss of CD4⁺ lymphocytes exceeds replenishment, immunoincompetence and disease occurs.

In Fig. 2 it is suggested that, in the early phase of HIV infection, ADCC-sensitized macrophages promptly recognize and destroy virus-producing cells; in the presence of sufficiently high levels of functional antiviral antibodies (RTI and NA), this would protect the immune system from extension of the cytopathic effects of the virus and the immunopathic effects of the viral proteins; this would permit the level of CD4⁺ lymphocytes to be maintained by replenishment. In the later phase, however, ADCC-sensitized macrophages destroy antigen-coated CD4⁺ cells as well as virus-producing cells and, in the process, self-destruct (with concomitant degradation of ADCC) at a rate that exceeds replenishment of both CD4⁺ cells and ADCC antibody. This suggests that cytotoxic antibodies (via ADCC) and/or cytotoxic T cells are important in both protection and pathogenesis depending upon the phase of infection and the cells destroyed: in the early phase primarily virus-infected cells, and in the later phase both virus-infected cells and cells coated with antigen (either antigen-producing cells or cells that have captured circulating

antigen on their surface).

If the persistence or decline of immunoprotective factors (anti-p24 and RTI) (Fig. 1) determines whether or not AIDS develops, then it may be possible, using an appropriately formulated immunogen in serologically positive individuals, to stimulate an anamnestic response early enough to enhance and prolong, perhaps indefinitely, the continued presence of any immunoprotective factors induced by the initial response to infection.

Intervention

To enhance and prolong the presence of such protective factors as are postulated to be induced by the initial immune response to infection (Fig. 1), a suitably potent non-infectious immunogen is needed to be tested by administration to serologically positive virus carriers. An immunogen for this purpose could be prepared from inactivated whole virus, either wild-type or replication-defective, or from a mixture of relevant antigens (subunits or recombinant DNA-produced proteins) incorporated in a potent immunological adjuvant. An appropriate adjuvant would heighten and prolong the immune response and permit the use of smaller quantities of antigen than would otherwise be required; it would also broaden the specificity of the immune response, especially in previously sensitized subjects.²³⁻²⁶

In the preparation of a whole-virus

immunogen, infectivity should be destroyed (by disrupting the function of viral nucleic acid as well as any residual cellular nucleic acid²⁷) while retaining the immunogenicity of the relevant protective antigens. Principles for achieving this have been previously described²⁸. For example, formalin has been used successfully under conditions that result in a linear rate of virus inactivation for a period of time sufficient to provide a wide margin of safety in respect to destruction of virus infectivity with retention of antigenicity; destruction of virus infectivity may also be achieved by use of psoralen, β -propiolactone or heat, as well as by formalin alone or in any combination²⁷. A further safeguard is provided by the additional requirement for demonstrating consistency in the preparation of successive batches free of detectable virus²⁸. Ultimately a replication-defective virus might be developed for use as a noninfectious immunogen.

Immunogens composed of noninfectious whole virus and of mixtures of recombinant DNA-produced proteins or subunits should be compared for immunogenicity to determine which are required to prevent (1) pathology (post-exposure) and (2) infection (pre-exposure). For prevention of disease by post-exposure immunization it would be sufficient for the immunogen to match antigenically the infecting type of HIV in the patient to be treated. For prevention of infection by pre-exposure immunization it would be necessary that the immunogen include all of the protective antigens and antigenic variants capable of causing infection in uninfected individuals. For prevention of disease in those who are serologically positive and with an intact immune system, one dose of an appropriately designed post-exposure immunogen should be sufficient to induce an anamnestic response. (Hyperimmune serum prepared in this way could be used to study the usefulness of passive immunotherapy in ARC and AIDS.) Prevention of infection using an appropriately designed pre-exposure immunogen in those who are serologically negative would require two or more doses to induce an adequate primary and secondary-type response.

It is evident that different strategies are needed to prevent infection (with HIV) or to prevent disease (AIDS). A strategy for the prevention of HIV infection using a suitable pre-exposure immunogen can be patterned on those used or contemplated for the prevention of such infections as influenza, poliomyelitis, rabies, hepatitis B, and herpes simplex, each of which has its own manifestations and requirements (Tables 1 and 2).

For the prevention of influenza, the immunogen must be administered pre-exposure, because the incubation period is short (Table 1) and protective antibody

Table 1 Characteristic effects of aetiological agents of six viral diseases

Viral disease	Incubation period	Clinical course	Cytopathicity	Latency	Immunopathicity
Influenza	days	acute	+	0	0
Poliomyelitis	days/weeks	acute	+	0	0
Rabies	weeks	acute	+	0	0
Hepatitis B	weeks/months	chronic	0	+	+
Herpes	weeks/years	recurrent	+	+	+
AIDS	months/years	progressive	+	+	+

+, occurs; 0, does not occur.

Table 2 Preventive effect of immunogen administration in six viral diseases

Viral disease	Pre-exposure			Post-exposure		
	Infection	Viral cytopathicity	Immunopathicity	Infection	Viral cytopathicity	Immunopathicity
Influenza	+	+	—	0	0	—
Poliomyelitis	+	+	—	0	0	—
Rabies	+	+	—	0	+	—
Hepatitis B	+	—	+	0	—	+
Herpes	+	+	+	0	+	+
AIDS	+	+	+	0	+	+

+, preventive; 0, not preventive; —, not applicable.

*Hypothetical.

must be present at the time of infection at the portal of entry, which is also the site of infection and of pathology²⁵. In poliomyelitis, with an incubation period of approximately one to two weeks, the immunogen must also be administered before exposure. The portal of entry for poliovirus is the oropharynx; the major site of viral multiplication is the intestinal tract and of pathology is the central nervous system (CNS)²⁵. The pathology of poliomyelitis can be prevented by the presence of neutralizing antibody in the bloodstream to block access of virus to the spinal cord and medulla. For prevention of paralysis in poliomyelitis, in contrast to influenza (a disease of shorter incubation period), it is not necessary for serum antibody to be present at the time of exposure to virus (in previously immunized or infected individuals) because an anamnestic response will occur early enough to prevent paralysis²⁹⁻³¹. In hepatitis B, with its relatively long incubation period, the pattern is similar to poliomyelitis in that infection post-vaccination induces an anamnestic-type response even if antibody is not detectable at the time of exposure. Thus for durable immunity both to paralytic poliomyelitis²⁹ and hepatitis B³² the induction merely of persistent immunological memory alone is sufficient.

Latent viral infection

For viruses that establish latency, it is necessary to prevent the initial cellular infection by induction of immunity before exposure. Once latency becomes established, then the immunopathic and cytopathic effects of the virus can develop (Table 1). Thus, to prevent the establishment of latency it would be necessary to administer a hepatitis B virus vaccine before exposure, to prevent the initial infection. The same would be required for herpes simplex virus and HIV infections, each of which possesses the capacity for latency.

The strikingly long incubation period between infection and onset of clinical manifestations in AIDS, as compared to the other diseases (Table 1), suggests that post-exposure intervention might be effective, as in rabies (cytopathicity) and hepatitis B (immunopathicity)^{33,34}, in

which it is possible to prevent pathology during a short period of time after as well as before exposure to the respective viruses (Table 2). Similarly, in herpes, post-exposure intervention might suppress or prevent the subsequent occurrence of cytopathicity and/or immunopathicity.

In AIDS the major pathological defect may be attributable to the immunoreactive viral proteins expressed by the viral genome (immunopathic) as well as to virus replication (cytopathic). Therefore, to prevent the development of AIDS by post-exposure intervention in an HIV-infected individual, it would be necessary to administer a suitably designed immunogen to prevent or suppress the production of viral proteins (antigenaemia) as well as of virus (viraemia).

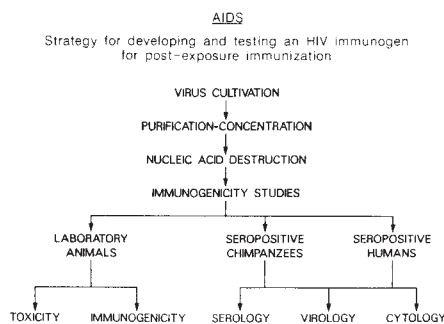
An analogy may be drawn between the forms of intervention required to prevent the immunopathic consequences in AIDS and in experimentally induced allergic encephalomyelitis (EAE) in animals. EAE is induced in laboratory animals by injection of myelin basic protein (MBP) emulsified in complete Freund's adjuvant (CFA) (which contains mycobacteria). EAE so induced is preventable by the administration of MBP emulsified in incomplete Freund's adjuvant (IFA) (which does not contain mycobacteria) either before or after the administration of MBP in CFA; prevention of EAE resulting from the inoculation of MBP in CFA can be effected if MBP in IFA is administered early enough in the period of incubation before disease onset³⁵, as is also proposed for the prevention of AIDS. The objective in preventing EAE in this way is to induce blocking antibody to prevent the development of the immunopathology of a delayed hypersensitivity type³⁶. The objective in preventing AIDS by post-exposure immunization is to enhance and prolong the presence of protective factors such as ADCC, RTI and NA, and possibly other protective antibodies as well.

The comparative observations regarding the immunopathological consequences of inoculating MBP emulsified in CFA versus IFA are of special interest in that they reveal the different effects of an immunogen, either in the induction of an immunopathologic disease (intolerance)

or its prevention (tolerance), depending upon the adjuvant used. Thus, the characteristics of the adjuvant chosen to enhance the protective immune response to HIV proteins will be critically important.

In HIV-seropositive human subjects, whether asymptomatic or symptomatic, the functional integrity of the immune system in respect to humoral anamnestic responsiveness can be tested by administering a commonly used vaccine; similarly the presence of cell-mediated immunity can be revealed by a test for delayed hypersensitivity responsiveness. Serum samples collected before and after administration of the HIV immunogen would reveal any changes induced in the antibody profile as determined by *in vitro* assays, including titration for NA, RTI and ADCC; and cell-mediated immunity would be revealed by antigenic stimulation of cells *in vitro* or by the size of delayed-type dermal hypersensitivity reactions. Over the longer term, studies of CD4⁺ lymphoid cell levels and functions would need to be carried out as well as tests for level of detectable virus in blood and, by implication, for contagiousness. Based on these findings, further investigations of an appropriate immunogen for immunoprophylaxis could be undertaken in serologically negative individuals as well as in appropriate animal model systems.

Animal models presently exist in chimpanzees and rhesus monkeys infected with HIV or simian T-lymphotropic virus III (STLV-III)³⁷, respectively, to test for the effects of an immunogen administered post-exposure, in terms of anamnestic response and virus carriage, and to be certain that unexpected adverse side effects do not occur in such animals and, therefore, presumably would not occur in humans. But the ability of immunization to

**Fig. 3**

prevent disease can only be answered by studies in human subjects who are already infected. Toward this end, appropriately designed immunogens are required to be tested in serologically positive (HIV) healthy individuals who are at risk of developing AIDS and who are also a source of contagion to others. A strategy for the development and testing of an HIV immunogen for post-exposure immunization is outlined in Fig. 3.

Perspective

As discussed above, a clue to strategies for immunological intervention in AIDS comes from its long incubation period compared to the other diseases cited in Table 1 and from the differences in the strategies needed for the prevention of each (Table 2). A further clue is found in the observation that the presence of anti-p24 and RTI antibodies correlate significantly with relatively slow disease progression in subjects infected with HIV. Thus, based on the foregoing observations and reasoning it might be possible, by the administration of an appropriately formulated immunogen, to prevent the immunodeficiency syndrome (AIDS) in HIV-infected individuals. If this can be achieved a further question to be answered concerns the extent of the interval into the incubation period, by passive or active means, during which this might be effective. Studies for immunoresponsiveness would need to be carried out over the entire spectrum from asymptomatic to symptomatic individuals. Those whose immunodeficiency is too far advanced to respond fully to active immunization may benefit from passive immunization. A study of the effectiveness of passive immunization, involving patients in various stages of disease, could be designed using immunoglobulins from pooled serum of hyperimmunized asymptomatic seropositives. If, in this way, the degree of immunodeficiency can be reduced to any significant extent, then it would be of interest to test the possible effectiveness of passive/active immunization. If progressive immunoincompetence can be arrested or moderated in this way, it will then be necessary to determine whether or not CNS disease³⁸ can similarly be controlled.

As to potential risks associated with administering an HIV immunogen to

in CD4⁺ lymphocytes). But the anticipated benefit of an HIV immunogen in those previously sensitized by HIV infection is based on the expectation that a significantly protective anamnestic response will occur in such individuals.

The studies here proposed have as their primary purpose the development of a means of the prevention of progressive immunological disease (AIDS), in healthy HIV carriers and, if possible, to arrest progression in symptomatic carriers whose disease has not yet advanced too far. The objective of the approach proposed is to destroy virus- and viral antigen-producing cells by the induction of the immune system's cytotoxic mechanisms known to rid the host of virus and virus-producing cells. If successful, HIV then would persist only in a proviral form in infected cells; if such cells are induced by antigenic stimulation they would promptly become targets of the cytotoxic immune response and be eliminated as foci of infection. The strategy proposed for immunizing seropositive HIV carriers would be tested for its effect in reducing the risk of disease in treated as compared to untreated individuals and in reducing contagiousness to others.

From an epidemiological perspective the prospects for the control of AIDS by immunization of seropositive individuals — if this proved to be successful in preventing the development of disease in those who are asymptomatic — would have a greater and more rapid impact in reducing HIV morbidity and mortality than would an immunization strategy directed at the seronegative population. If immunization of seropositive virus carriers would also reduce their contagiousness then the virus reservoir in the population would be rapidly reduced, as would the frequency of newly acquired infections. The prospect of producing such effects would increase the frequency of blood-testing, for reasons of self-interest, without the need for coercion. This would have the further advantage of accelerating case-finding and early treatment of those so identified.

In the absence of precise knowledge as to which of the many antigens of the virus induce a protective effect, a non-infectious form of the whole virion is proposed for immunizing seropositive individuals. Concerns that have been raised, in general regarding the use, in seronegative individuals, of the whole HIV virion (including its inactivated nucleic acids), would not be relevant in seropositives in whom the proviral HIV genome is already incorporated in host DNA.

Another important consideration concerns the maximally effective means for presenting the immunogen. Every possible advantage should be taken to maximize the immune response, including the use of the most potent immunological

adjuvants. Strategically, the practice of immunization of both asymptomatic and symptomatic HIV carriers, as here proposed, is akin to the administration of drug-therapy intended to reduce the virus burden in the host and to reduce both morbidity and mortality, and contagiousness.

The purpose of this paper is to invite comment and discussion and to elicit suggestions and data relevant to the hypothesis and strategy proposed for the control, by post-exposure immunization, of this rapidly spreading infection and its associated diseases. □

I thank Drs F. Barre-Sinoussi, J. C. Chermann, Melvin Cohn, Donald Francis, Murray Gardner, Paul Luciw, David Pauza, Niel Pedersen, Jonathan Salk, Peter Salk, Maurizio Zanetti for discussions and suggestions, and Barbara Robinson for preparing the manuscript.

1. Gluckman J.C., Klatzmann D. & Montagnier L. *A. Rev. Immun.* **4**, 97-117 (1986).
2. Fauci A.S. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9278-9283 (1986).
3. Barre-Sinoussi, F. *et al. Science* **220**, 868-871 (1983).
4. Gallo, R.C. *et al. Science* **224**, 500-503 (1984).
5. Levy J.A., Hoffman, A.D., Kramer, S.M., Landis, J.A. & Shimabukuro, J.M. *Science* **225**, 840-842 (1984).
6. Bloom, B.R. & Rager-Zisma, B. in *Viral Immunology and Immunopathology*. (ed. Notkins A.L.) 113-136. (Academic, New York, 1975).
7. Pfizenmaier, K., Jung, H., Starsinski-Powitz, A., Rollinghoff, M. & Wagner H. *J. Immun.* **119**, 939-944 (1977).
8. Edgington, T.S. & Chisari, F.V. *Am J Med Sci* **270**, 213-227 (1975).
9. Peterson, J.M. Dienstag, J.L. & Purcell, R.H. in *Viral Immunology and Immunopathology* (ed. Notkins, A.L.) 213-235 (Academic, New York, 1975).
10. Adams, C.S., Crawford, T.B., Banks, K.L., McGuire, T.C. & Perryman, L.E. *Infect Immun.* **28** 421-427 (1980).
11. McGuire, T.C. *Concepts in Viral Pathogenesis*. (eds. Notkins, A.L., Oldstone, M.B.A.) 225-230 (Springer, New York, 1984).
12. Klatzmann, D. & Gluckman, J.C. *Immun. Today* **7**, 291-296 (1986).
13. Wong-Staal, F. & Gallo, R.C. *Nature* **317**, 395-403 (1985).
14. McDougal, J.S. *et al. J Immun.* **135**, 3151-3162 (1985).
15. Folks, T. *et al. J Immun.* **136**, 4049-4053 (1986).
16. Fauci, A.S. *Proc Soc Acad Sci U.S.A.* **83**, 9278-9283 (1986).
17. Polk, B.F. *et al. N. Engl. J. Med.* **316**, 61-66 (1987).
18. Weber, J.N. *et al. Lancet* **i**, 119-123 (1987).
19. Steimer, K.S. *et al. Virology* **150**, 283-290 (1986).
20. Steimer, K.S. *et al. Vaccines* **87** (eds. Chanock, R., Brown, F., Levner, R. & Ginsberg, H.) (Cold Spring Harbor Laboratory, New York, in the press).
21. Laurence, J., Saunders, A. & Kulkosky, J. *Science* **235**, 1501-1504 (1987).
22. Klatzmann, D. *et al. Science* **225**, 59-63 (1984).
23. Salk, J. *et al. J Am med Assoc* **515**, 1169-1175 (1953).
24. Meiklejohn, G. *J Am med Assoc* **179**, 594-597 (1962).
25. Salk, J. & Salk, D. *Science* **195**, 834-847 (1977).
26. Allison, A.C. & Byars, N.E. *J Immun. Meth.* **95**, 157-168 (1986).
27. Hanson, C.V. *Medical Virology II* 45-79 (eds. de la Maza, L.M., Peterson, E.M.) (Elsevier, New York 1983).
28. Salk, J. & Gori, J.B. *Ann N.Y. Acad Sci* **83**, 609-637 (1960).
29. Salk, J. *Ann N.Y. Acad Sci* **61**, 1023-1036 (1955).
30. Salk, J. *Lancet* **ii**, 715-723 (1960).
31. Salk, D., van Wessel, A.L. & Salk, J. *Lancet* **ii**, 1317-1321 (1984).
32. Hadler, S.C. *et al. New Engl J. Med.* **315**, 209-214 (1986).
33. Francis, D.P. *et al. Ann. int Med* **97**, 362-366 (1982).
34. Szmuness, W. *et al. New Engl J Med* **303**, 833-841 (1980).
35. Salk, J., Romine, J.S., Westall, F.C. & Wiederholt, W.C. in *The suppression of experimental allergic encephalomyelitis and multiple sclerosis* (eds. Davison, A.N. & Cuzner, M.L.) **141-156** (Academic, London, 1980).
36. Salk, J. *Ann N.Y. Acad Sci* **162**, 365-380 (1969).
37. Letvin, N.L. *et al. Science* **230**, 71-73 (1985).
38. Koenig, S. *et al. Science* **233**, 1089-1093 (1986).

Jonas Salk is at the Salk Institute for Biological Studies, PO Box 85800, San Diego, California 92138-9216, USA.

Table 3 Clinical effects of antigenic stimulation in HIV carriers

Antigenic stimulation	Anticipated effects	
	Pathogenic	Protective
Infections	+	-
Non-HIV immunogens	+	-
HIV immunogens	-	+

HIV-infected individuals, Table 3 shows the anticipated effects, pathogenic or protective, of antigenic stimulation by either superinfections or by non-HIV immunogens as compared to an appropriately formulated HIV immunogen presented in a suitable adjuvant. *A priori*, a protective effect would not be anticipated from superinfections or from antigenic stimulation with a non-HIV immunogen, both of which can be expected to have a pathogenic effect (induction of virus expression

Variations in mode of formation and temperature of oceanic deep waters over the past 125,000 years

L. D. Labeyrie*, J. C. Duplessy* & P. L. Blanc†

* Centre des Faibles Radioactivites, Laboratoire Mixte CNRS/CEA Parc du CNRS 91190 Gif/Yvette, France

† CEA/IPSN/DPT CEN/FAR BP no. 6-92265 Fontenay aux Roses Cedex, France

Marked changes in the deep-water temperature of the different oceanic basins during the past 125 kyr are revealed by a detailed comparison of the oxygen isotopic composition of benthic foraminifera from sediment cores taken from the Norwegian Sea, and the Pacific, Atlantic and Southern oceans. The last interglacial deep-water temperatures were similar to the modern ones. At the 5e/5d transition, around 115 kyr BP, the deep-water temperatures dropped below approximately 0 °C in the Pacific and southern Indian oceans. In the Atlantic Ocean, the deep-water temperature dropped in 2 successive steps at ~115 and 75 kyr BP. The global ocean deep-water temperatures were in the 0 to -1 °C range for the whole of the last ice age (isotope stages 4-2).

SINCE Emiliani's pioneer work¹, a large number of studies dealing with oceanic palaeoclimatology have used the oxygen isotope record of foraminifera from deep-sea sediments as a tool for stratigraphic correlations. Until now, the interpretation of these variations in terms of hydrology was limited by the inability to differentiate directly between the relative effects on the foraminiferal $\delta^{18}\text{O}$ of the temperature and the $\delta^{18}\text{O}$ of the water in which the calcite was precipitated^{2,3}. Whereas the relationship between temperature of growth and oxygen isotope fractionation has been determined⁴⁻⁶, the variations of the water $\delta^{18}\text{O}$, linked to the amount of continental ice, are unknown. Here we present the demonstration that the Norwegian Sea has been the location of active formation of deep water throughout the last interglacial and early glacial period. As a consequence, during that period, the Norwegian Sea deep-water temperature was constant at approximately -1 °C, and the benthic foraminiferal $\delta^{18}\text{O}$ variations recorded solely those of the water $\delta^{18}\text{O}$. The global sea water $\delta^{18}\text{O}$ record has been extended to the whole last climatic cycle by patching the Norwegian Sea benthic $\delta^{18}\text{O}$ variations with those of the Pacific Ocean core V 19-30 (ref. 7). Using this record, which closely fits the sea-level record, we may calculate the fraction of the benthic foraminiferal $\delta^{18}\text{O}$ signal due to deep-water temperature changes in the different oceanic basins.

Global deep-water circulation

In the modern ocean, deep waters are composed mainly of a mixture of relatively warm (2 °C) and saline (over 34.9‰) North Atlantic deep water (NADW) with cold (<0 °C) and less saline (<34.7‰) Antarctic bottom water (AABW). Deep-water formation starts in winter in high-latitude surface waters. Seawater density increases continuously during cooling until the freezing temperature is reached (~-1.9 °C). Salt is rejected during sea-ice production, further increasing the density of sea water. This process may lead to cold (~-1 °C) bottom water formation in favourable locations. This is the case in various places around Antarctica, especially in the Weddell Sea⁸, but also along the continental shelf of the Labrador Sea⁹ and in the Norwegian Sea. Compared to these various water masses, NADW is exceptionally warm and saline. This is due to the peculiar location of the modern North Atlantic polar front during summer, approximately south-north from 60 to 70° N across the Norwegian Sea, which allows surface penetration of saline subtropical water far to the north. The strong winter cooling and sea-ice formation produces a large volume of dense water. This cold (~-1 °C) Norwegian Sea bottom water mixes with warm

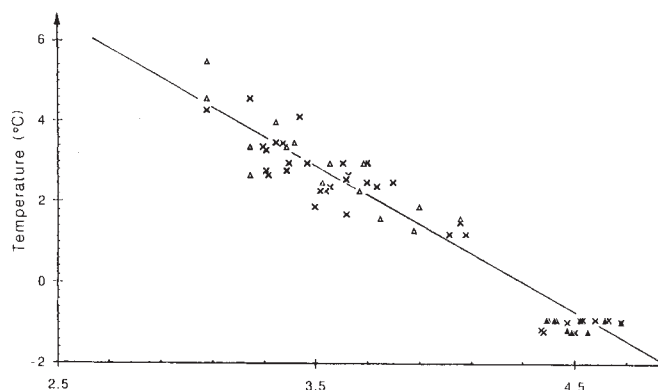


Fig. 1 Modern bottom-water temperature at different core locations³³⁻³⁶, expressed as a function of the oxygen isotope fractionation as $\delta^{18}\text{O}_{\text{foraminifera}} - \delta^{18}\text{O}_{\text{water}}$ (both versus PDB), for benthic foraminifera extracted from Holocene core tops. The bottom water $\delta^{18}\text{O}$ are calculated¹⁴ using bottom-water salinities³³⁻³⁶. Different foraminiferal species have been analysed. Measured $\delta^{18}\text{O}$ are corrected for specific fractionation¹³, by +0.64‰ for * *Cibicides wuellerstorfi*, by +0.36‰ for † *Oridorsalis tener* and no corrections for ‡ *Uvigerina peregrina*. We have also reported the line defined by Shackleton⁴ as representing equilibrium isotope fractionation, derived from the high-temperature experiments of O'Neil *et al.*³⁷.

(3-6 °C) and saline (34.9-35 ‰)¹⁰ North Atlantic intermediate water while overflowing through the Denmark and Faeroe sills to the North Atlantic basins. This gives the major component of the NADW.

This hydrological situation has not been constant over the Pleistocene climatic cycles. During the last glaciation, the mean position of the polar front shifted to a more latitudinal orientation far south of the sills¹¹. This reduced both the warm and saline water contribution to the Norwegian Sea and also the mixing of this warm water with Norwegian Sea bottom water at the sills. Furthermore, other sources of cold deep water developed, probably over the shallow parts of northern Atlantic Ocean, near the ice margins, if the modern Weddell Sea is taken as analogue^{12,22}. These added effects explain the colder deep water observed in the north Atlantic Ocean during the last glaciation¹². Accordingly, the amplitude of the glacial/interglacial changes in benthic foraminiferal $\delta^{18}\text{O}$ is significantly higher in the Atlantic than the Pacific oceans^{7,12}.

Calibration of benthic foraminiferal $\delta^{18}\text{O}$

Two assumptions are necessary to interpret the isotopic records in quantitative terms. First, the $\delta^{18}\text{O}$ of the different species of benthic foraminifera, must record quantitatively the water temperature, taking into account the foraminiferal specific fractionation factors^{4,13}. Second, the changes in seawater $\delta^{18}\text{O}$ have to be separated from the temperature effect in the benthic foraminiferal $\delta^{18}\text{O}$ signal. The first assumption has been tested in different Holocene core tops, by comparing the oxygen isotope fractionation ($\delta^{18}\text{O}_{\text{carbonate}} - \delta^{18}\text{O}_{\text{water}}$) of three benthic foraminiferal species: *Cibicides wuellerstorfi*, *Oridorsalis tener* and *Uvigerina peregrina*, with the deep water temperatures as observed today (Fig. 1). These values are corrected for specific fractionation (ref. 13 and Table 1). Using observed salinities¹⁴ $\delta^{18}\text{O}_{\text{water}}$ is calculated. We have also plotted in Fig. 1 the equilibrium relationship between isotope fractionation and temperature as estimated by Shackleton for the species *Uvigerina*⁴. Mean difference between the deep-water temperatures estimated for the Holocene and the modern measured values is less than 0.1 °C, with a standard deviation of 0.5 °C, which justifies the use of Shackleton's equation to calculate past deep-water temperature changes. The three different species analysed appear similarly reliable.

Norwegian Sea $\delta^{18}\text{O}$ record

To estimate the mean $\delta^{18}\text{O}$ changes of sea water during the last ice age, we propose to use benthic foraminiferal $\delta^{18}\text{O}$ records from areas where bottom-water temperature changes have been a minimum. The Norwegian Sea is a good candidate for such characteristics, as its waters below 1,000 m are already at -1 °C (± 0.2 °C). This low temperature is regulated by the intense cooling of the surface water, sea-ice formation, and deep convection, during winter. As the Norwegian Sea surface water was covered with ice (sea ice or ice shelves) for most of the last glaciation¹⁵, with more or less localized ice-free areas in the southern part, at least between isotope stage 5d and the beginning of stage 3 (ref. 16), it is safe to admit that the bottom-water temperature remained at -1 °C ± 0.5 °C as long as energy loss at the surface was large enough to allow deep-water formation. In such a constant temperature environment, the foraminiferal $\delta^{18}\text{O}$ would record only the seawater $\delta^{18}\text{O}$ changes.

In addition, benthic foraminiferal $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ are useful tools to recognize if and where cold bottom-water formation occurs. In a set of contemporaneous samples from different

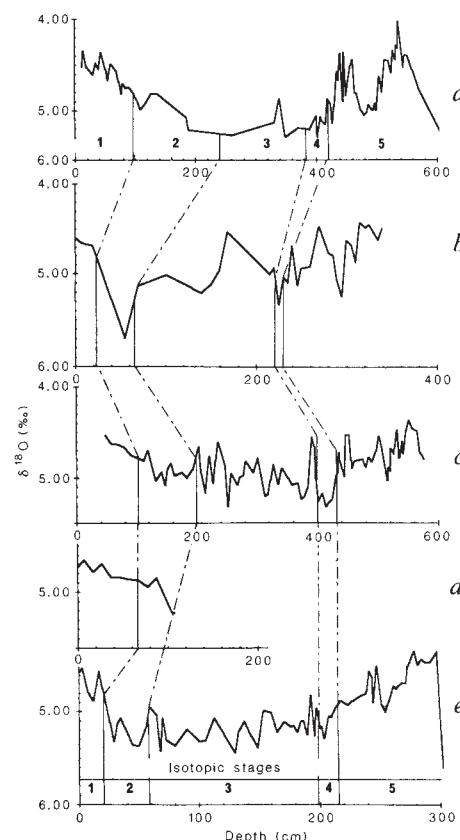


Fig. 2 $\delta^{18}\text{O}$ versus PDB of benthic foraminifera, in the *Uvigerina* scale¹³, expressed against depth in the different cores analysed for the study²⁰. a, Core V27-60 (part of the analyses have been kindly provided by N. Shackleton); b, core V27-86; c, core V28-38; d, core CH77-07; e, core K-11 (resampled for this study).

areas, the highest $\delta^{18}\text{O}$ corresponds to the coldest temperature and the highest $\delta^{13}\text{C}$ to the better ventilated water¹⁷. We have therefore analysed $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ in planktonic (*N. pachyderma* s.) and benthic (*C. wuellerstorfi* and *O. tener*) foraminifera from five Norwegian Sea cores (K-11, V28-38, V27-86, V27-60, CH77-07). Core locations are listed in Table 1. These cores have been studied previously for their faunal and sedimentological

Table 1 Location of the cores and comparison between of the Holocene benthic foraminifera $\delta^{18}\text{O}$, bottom-water $\delta^{18}\text{O}$ and temperature

Core	Latitude	Longitude	Water depth (m)	Depth in the core (cm)	Foraminifera $\delta^{18}\text{O}$	$\delta^{18}\text{O}$ versus PDB	Temp calc. (°C)	Modern temp. (°C)	Temp. diff. (°C)
V27-86	66°36' N	01°07' E	2,900	1	4.59*	0	-1.10	-0.90	-0.20
					4.62†	0	-1.20	-0.90	-0.30
V27-60	72°11' N	08°35' E	2,525	12	4.52*	0	-0.85	-0.90	0.00
					4.52†	0	-0.85	-0.90	0.00
V28-38	69°23' N	04°24' W	3,411	48	4.42*	0	-0.51	-0.90	0.40
					4.4†	0	-0.44	-0.90	0.50
K11	71°47' N	01°36' E	2,900	5	4.47*	0	-0.65	-0.95	0.30
					4.68†	0	-1.41	-0.95	-0.46
CH73-139c	54°38.2' N	16°21.3' W	2,209	0	3.31*	0	3.50	3.30	0.20
MD73-025	43°49' S	51°19' E	3,284	0	3.62*	-0.4	0.91	1.20	-0.29
MD84-527	43°49' S	51°19' E	3,269	5	3.73*	-0.4	0.52	1.20	-0.68
M12-392	25°10' N	16°50' W	2,573	0.5	3.3*	0	3.54	3.40	0.14
					3.25‡	0	3.72	3.40	0.32
V19-30	03°21' S	83°21' W	3,091	0	3.48‡	-0.4	1.41	1.30	0.10
CH77-07	66°36' N	10°31' W	1,487	0	4.70‡	0	-1.45	-1.10	-0.35

We report the depth of the sample; the foraminifera $\delta^{18}\text{O}$ corrected for specific fractionation¹³ (+0.64‰ for *Cibicides wuellerstorfi* (*), +0.36‰ for *Oridorsalis tener* (†), and no corrections for *Uvigerina peregrina* (‡)); the bottom water $\delta^{18}\text{O}$ calculated¹⁴ using bottom-water salinities³³⁻³⁶; the bottom-water temperature calculated from the foraminiferal $\delta^{18}\text{O}$ as in Fig. 1; the estimated modern bottom-water temperature at the core location³³⁻³⁶ and the difference between these two temperature estimates. Analyses of the core M12-392 have been taken in ref. 21 and of the core V19-30 in ref. 7.

characteristics and their stratigraphy is clearly established^{15,16,18}. Core K-11 has been resampled for this study, as earlier $\delta^{18}\text{O}$ analyses¹⁹ were based upon multispecific samples, and were therefore not reliable for quantitative interpretation. Complete analytical results will be presented in an accompanying paper²⁰. In Fig. 2 the benthic foraminiferal $\delta^{18}\text{O}$, corrected for specific fractionation is plotted against depth in the cores. Interpretation of these $\delta^{18}\text{O}$ records is complicated by the fact that during isotope stage 2 and part of stage 3, benthic foraminifera are rare or absent in the Norwegian Sea¹⁶. The few individuals picked for isotope analyses could be reworked, and this part of the record is dubious. It is especially true for cores V27-60 (Fig. 2a) and V27-86 (Fig. 2b). Cores V28-38 (Fig. 2c) and K-11 (Fig. 2e) contain enough *O. tener* to give a reliable $\delta^{18}\text{O}$ for stage 3. During isotope stages 1, 4 and 5, benthic foraminifera are abundant (both *O. tener* and *C. wuellerstorfi*), and detailed interpretation of the foraminiferal $\delta^{18}\text{O}$ signal is possible. Compared to typical benthic foraminiferal records from the Atlantic^{12,21} and Pacific⁷ oceans, large differences are apparent. Such differences are particularly significant during isotope stage 5. Figure 3a presents an expanded part of the benthic foraminiferal $\delta^{18}\text{O}$ record in core V28-38, during isotope stages 4 and 5, compared to the corresponding data in Pacific Ocean core V19-30 (ref. 7). Figure 3a shows that the benthic foraminiferal $\delta^{18}\text{O}$ in the Norwegian Sea is more positive than in the Pacific Ocean by $\sim 0.5\%$ over the period covering stages 4 to 5d, and by 1% during stage 5e. The Norwegian Sea $\delta^{18}\text{O}$ record has about the same isotope enrichment compared to other benthic foraminiferal records from the Atlantic and Indian oceans^{12,21,22}. This indicates that, at least over the whole of isotope stages 4 and 5, the Norwegian Sea bottom water was colder than in the other oceanic basins. If, as hypothesized¹⁵, salinity decreased in the Norwegian Sea at the first inception of the glacial, at ~ 115 kyr BP, the deep-water temperature should have been even colder to give the observed isotope difference. Yet, the isotope effect of the salinity decrease has not been larger than 0.2‰, as seawater temperature may not decrease below $\sim -1.7^\circ\text{C}$ at a depth of 2,500 m, due to the adiabatic warming of water while sinking to great depth.

The carbon isotope ratio of the benthic foraminiferal species *C. wuellerstorfi* may be used to identify the periods of formation of deep water, because it records the $\delta^{13}\text{C}$ of the total dissolved CO_2 of the sea water in which it grows, which is dependent upon deep-water ventilation¹⁷. We may compare (Fig. 3b) the *C. wuellerstorfi* $\delta^{13}\text{C}$ values of the Norwegian Sea core V28-38 with those of the north Atlantic Ocean core CH 73-139C and the Southern Ocean core MD84-527, located near the antarctic source of deep-water recirculation. The Norwegian Sea $\delta^{13}\text{C}$ record is significantly more positive than the other records, for all the period covered (130–60 kyr BP). While the increase in $\delta^{18}\text{O}$ of V28-38 between 400 and 430 cm (Fig. 3a) indicates without ambiguity glacial stage 4 (at ~ 60 –70 kyr BP), we do not observe the significant $\delta^{13}\text{C}$ decrease displayed in the other oceanic basins at the stage 5a–4 transition²². Thus, Norwegian Sea ventilation and cold deep-water formation were continuously active during isotope stages 5 and 4, in agreement with the rich and diversified benthic fauna of the period¹⁶. This result does not imply that the contribution of Norwegian Sea deep water to the global oceanic circulation represented a significant flux, but as long as deep water was forming, we may admit the stability of the Norwegian Sea deep-water temperature around -1°C , both during the interglacial stages 5e–5a, and during the first strong glaciation (isotope stage 4). The benthic foraminiferal isotope record may therefore be used during that period as a reference for the $\delta^{18}\text{O}$ changes in sea water with an accuracy of better than 0.2‰. The large $\delta^{18}\text{O}$ change presented by core V19-30 during the stage 5e–5d transition (Fig. 3a) corresponds therefore to a major cooling of deep Pacific Ocean water (by approximately 2°C), at the 5e–5d transition (~ 110 kyr BP).

During isotope stage 2 and most of stage 3, the benthic

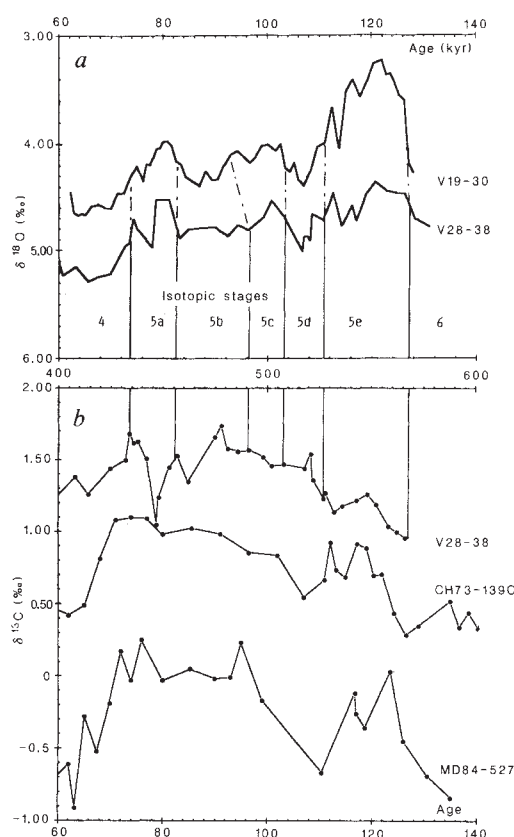


Fig. 3 a, Comparison between the $\delta^{18}\text{O}$ of benthic foraminifera in the 400- to 600-cm section of Norwegian Sea core V28-38(1) and the corresponding $\delta^{18}\text{O}$ record of Pacific Ocean core V19-30(2). b, Same comparison for the $\delta^{13}\text{C}$ record of *C. wuellerstorfi* in core V28-38(1) with the equivalent record for North Atlantic core CH73-139C¹²(2) and Southern Ocean core MD84-527²²(3).

Table 2 Control points for the expression of the Norwegian Sea $\delta^{18}\text{O}$ results in function of time

K11	V28-38	V2760	V27-86	CH77-07
(cm)	(cm)	(cm)	(cm)	(cm)
0	0	0	0	0
15	10.5	6	9	100
20	13	112	10.5	30
30	15	164	15	50
59	25	204	15	80
90	32.7	222	240	120
100	33.5	228	280	130
113	37	236	320	169
130	40.7	253	360	215
154	44.8	302	45	225
171	50.3	328	48.8	235
186	51.8	342	403	240
190	56.1	376	405	246
193	57.4	391	420	270
198	60.3	408	440	295
204	65	414	450	300
218	71	436	461	325
242	80	450	475	340
255	87	481	485	
270	99	502	500	
272	100.5	517	515	
274	117	528	560	
285	122	532	640	
300	127	541		
303	135	552		
		578		

The timescale is derived from Imbrie *et al.*²³ and Paterne *et al.*²⁴.

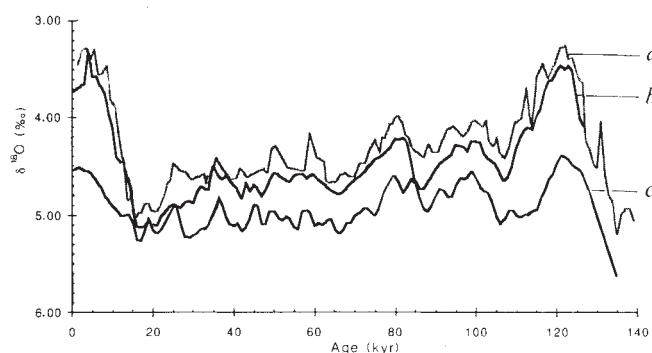


Fig. 4 Comparison over the last 140 kyr between the benthic foraminiferal $\delta^{18}\text{O}$ stack built for Norwegian Sea (a), the equivalent stack built for the Southern Ocean (b), and the *Uvigerina* $\delta^{18}\text{O}$ record of Shackleton *et al.*⁷ in core V19-30 (c).

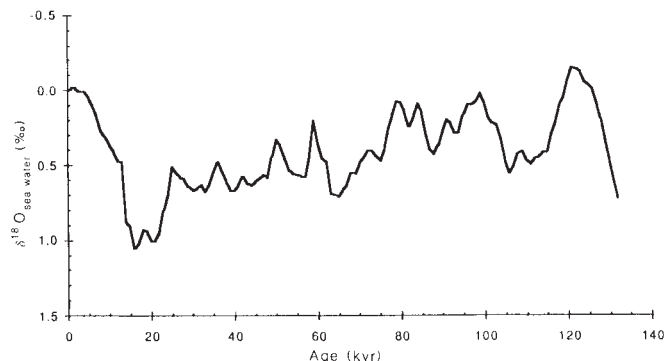


Fig. 5 Reconstructed changes in the mean seawater $\delta^{18}\text{O}$ for the last 130 kyr.

foraminiferal species *C. wuellerstorfi* is absent, except for a few reworked individuals, from the Norwegian Sea sediments. It is replaced by another species, *Oridorsalis tener*, presently abundant at depths greater than 2,900 m¹⁶. The $\delta^{13}\text{C}$ of this species does not vary linearly with the seawater $\delta^{13}\text{C}$ (ref. 20), and we may not use it to indicate whether deep water was actively forming. But its $\delta^{18}\text{O}$ was as positive during stage 3 (around +5‰) as that in the North Atlantic core CH73-139C¹², located in an area of cold deep-water formation during the glacial¹² period. This indicates that, contrary to earlier hypotheses^{16,19}, the Norwegian Sea deep water did not warm up significantly during that period.

The different cores studied have a great variability in sedimentation rates. Also we have built a continuous $\delta^{18}\text{O}$ record to allow detailed comparison of the Norwegian Sea signal with equivalent records from other oceanic basins, by stacking the benthic foraminiferal $\delta^{18}\text{O}$ analyses. For that stacking, ages have been calculated for the individual $\delta^{18}\text{O}$ values in each core from a common age scale referred to as the SPECMAC timescale²³ refined during isotope stage 3 with the chronostratigraphy of Paterne *et al.*²⁴ and during the last deglaciation by AMS C-14 dating of the foraminifera²⁵. Age-depth control points are listed in Table 2. They are defined using the stratigraphy of Streeter *et al.*¹⁶, Kellogg¹⁵, and Grousset and Duplessy¹⁸, based on sediment and micropalaeontological characteristics, and adjusted by detailed comparison between the $\delta^{18}\text{O}$ and $\delta^{13}\text{C}$ of the planktonic foraminiferal records of *N. pachyderma* in the different Norwegian Sea cores²⁰ with equivalent records in the north Atlantic Ocean²². Taking into account bioturbation differential mixing²⁶, we have made adjustments of a few cm to strengthen correlations between the benthic foraminiferal isotope records. Sedimentation rates are taken as constant between the control points. Once ages have been calculated for all individual $\delta^{18}\text{O}$ measurements, they have been stacked together. The stack record has been interpolated for each 0.33 kyr using a cubic spline, and smoothed using a 15-point least-squares quadratic filter (approximately 5 kyr window)²⁷.

Similarly, for the Southern Ocean, we have stacked the $\delta^{18}\text{O}$ analyses of *Nonion* sp. in core MD73-025 (ref. 12) and *Cibicides wuellerstorfi* in core MD 84-627 (ref. 28). This record is more detailed, but not significantly different from the previously published record of MD 73-25 (ref. 12). The reference for the Pacific Ocean is core V19-30 (ref. 7). Location of the cores is given in Table 1.

The resulting benthic foraminifera $\delta^{18}\text{O}$ records are presented in Fig. 4. The differences between records are due to the gradients in temperature and $\delta^{18}\text{O}$ of the deep water between the three basins. The covariance between the Southern and Pacific oceans records is logical, as the Southern Ocean site is located within circumpolar deep water, the major source of deep water for the Pacific Ocean. Most of the foraminiferal isotope difference, in

this case, may be attributed to the difference in water temperature (approximately 1 °C), due to the proximity of the Southern Ocean site from the source of AABW. We have already discussed the differences presented by the Norwegian Sea records during stage 5. It is clearly apparent from Fig. 4 that most of the deep-water hydrological changes occurred at the end of stage 5e (115 kyr BP) for the inception of the deep water glacial circulation, and at the stage 2-1 transition (10-14 kyr BP) for the return to interglacial conditions. A regression analysis between the $\delta^{18}\text{O}$ of core V19-30 and the $\delta^{18}\text{O}$ of the Norwegian Sea stack record for the 13-107 kyr period gives a slope of 1.05 (± 0.09) and a correlation coefficient of 0.78, highly significant taking into account the poor quality of the Norwegian Sea record during isotope stage 2 (13-25 kyr BP). Detailed adjustment between the two records of the small $\delta^{18}\text{O}$ changes (for example, 60 or 85 kyr BP) would strengthen that correlation. This close covariance justifies the hypothesis that the global conditions of the formation of deep water reached during isotope stage 5d did not change until the end of the glaciation. We may therefore admit that the major deep-water temperature changes occurred for the Pacific Ocean around 110-115 kyr BP and 10-15 kyr BP, and that the temperature did not vary noticeably within this interval.

Interoceanic water $\delta^{18}\text{O}$ gradients

To interpret in terms of temperature the $\delta^{18}\text{O}$ difference between Norwegian Sea and Pacific Ocean foraminiferal records, we have to estimate the seawater $\delta^{18}\text{O}$ gradient between the corresponding water masses. The Norwegian Sea $\delta^{18}\text{O}$ is today at +0.2‰ against SMOW (ref. 14), and deep Pacific waters -0.2‰, due to the effect of lower salinity AABW. The water $\delta^{18}\text{O}$ gradient is negligible between the Norwegian Sea and the north Atlantic deep waters¹⁴. During the glaciation, the southward migration of the north Atlantic polar front towards a more latitudinal position around 45° N¹¹ might have isolated a polar surface water of relatively lower salinity and $\delta^{18}\text{O}$. Both the Norwegian Sea and north Atlantic deep waters would have been similarly affected by this salinity decrease, as long as deep-water formation was occurring in both basins. With such a hypothesis, the isotope gradient between the Norwegian Sea and Pacific Ocean deep waters might have been reduced.

Because isotope stage 4 is a full glacial, with deep water still actively forming in the Norwegian Sea, it is a favourable period to test the possibilities of changes in seawater $\delta^{18}\text{O}$ gradients. The Norwegian Sea $\delta^{18}\text{O}$ of water at that time, calculated using the benthic foraminiferal $\delta^{18}\text{O}$ of +5.16‰, is +0.8‰ for a water temperature of -1 °C. Assuming the same isotope ratio for the deep water of the north Atlantic Ocean at the location of core CH 73-139C, the water temperature (derived from the benthic foraminiferal $\delta^{18}\text{O}$ of +4.64‰)¹² was +0.2 °C, as opposed to +2.5 °C now. This low value confirms that there was no more significant input of 'warm' NADW to the world ocean during isotope stage 4. Taking into account the further cooling by

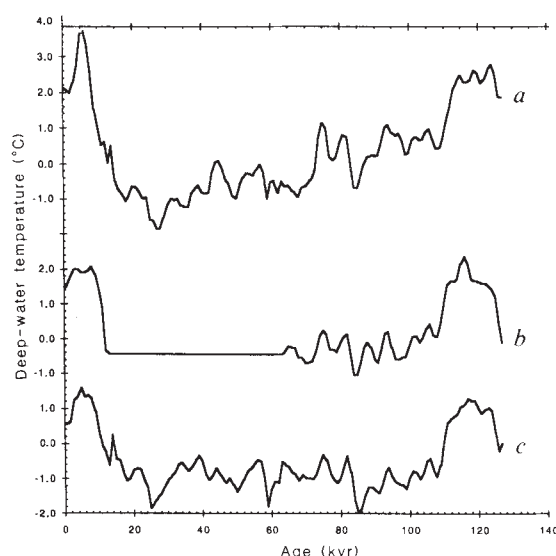


Fig. 6 Changes in the deep-water temperature over the last climatic cycle. *a*, Atlantic Ocean, using the benthic foraminiferal $\delta^{18}\text{O}$ from core M12-392 (ref. 21); *b*, equatorial Pacific ocean, using the benthic foraminiferal $\delta^{18}\text{O}$ of core V19-30 (ref. 7); *c*, south Indian Ocean, using the stacked $\delta^{18}\text{O}$ record of cores MD73-025 and MD84-527. The estimated error is approximately $\pm 0.7^\circ\text{C}$ and individual peaks may be due to slight mismatches in the time scales of the benthic records compared to the water reference $\delta^{18}\text{O}$ record.

mixing with Antarctic bottom water, the maximum temperature of the deep waters of the Pacific and Southern oceans must have been below 0°C . The benthic foraminiferal $\delta^{18}\text{O}$ value of $+4.66\%$ for the Pacific Ocean core V19-30 during stage 4 (Fig. 4c) gives a water $\delta^{18}\text{O}$ value of $+0.55\%$ for a 0°C temperature, and $+0.3\%$ for -1°C . These values indicate that the Norwegian Sea/Pacific Ocean isotope gradient was in the range 0.25 to 0.5‰ not significantly different from the modern gradient. In the absence of other constraints, we take, for our reconstruction, an isotope gradient of 0.4‰, knowing that we may have overestimated it slightly.

Past variations of $\delta^{18}\text{O}$ and temperature

Because the water temperature in the deep Pacific Ocean probably remained constant during the glaciation¹², the isotope record of core V19-30 may be used there as reference for the seawater $\delta^{18}\text{O}$ changes. Therefore, we have derived the mean global $\delta^{18}\text{O}$ record of sea water by piecing together the Norwegian Sea $\delta^{18}\text{O}$ record for the interglacial periods (0–12 and 65–135 kyr BP), and the $\delta^{18}\text{O}$ record of core V19-30 (corrected by $+0.57\%$, the mean difference between the Norwegian Sea and V19-30 records over the period 65–105 kyr BP), for the glacial period (12–65 kyr BP). This record has been corrected by -4.53% to take into account the oxygen isotope fractionation of the benthic foraminifera at the core top (Fig. 4). The resulting changes in mean seawater $\delta^{18}\text{O}$ for the last climatic cycle are reported in Fig. 5. All the data will be detailed in ref. 20.

We now have the elements to reconstruct the changes in water temperature for this period in the different oceanic basins by subtracting the reference seawater $\delta^{18}\text{O}$ records from each benthic foraminiferal record. Results of the calculations are reported (see Fig. 6). The deep-water temperature changes are estimated in the Atlantic Ocean (Fig. 6a) using the benthic foraminiferal $\delta^{18}\text{O}$ record of core M12-392 (ref. 21). The deep Pacific Ocean is represented by core V19-30 (Fig. 6b). The southern Indian Ocean temperature reconstruction (Fig. 6c) is based upon the stacked benthic foraminiferal records of cores MD73-025 and MD84-527 (Fig. 4).

To interpret these temperature records, we must keep in mind the different hypotheses involved, which limit the precision of the reconstructions to approximately $\pm 0.7^\circ\text{C}$, assuming a mean error of $\pm 0.1\%$ for each of the records. In the Pacific core V19-30 (Fig. 6b) and the southern Indian Ocean (Fig. 6c), the cooling to glacial temperatures around or below 0°C , occurred within a few thousand years, between 115 and 110 kyr BP. This also dates the first major change in north Atlantic hydrology^{13,29}. However, in the Atlantic Ocean (Fig. 6a), a second cooling phase was associated with the stage 5a–4 transition, so that the temperature dropped first by 2.5°C at 115 kyr BP, and 1°C at 70 kyr BP. This corresponds to the southward migration of the north Atlantic polar front during that period¹¹. The deep-water temperature remained in the 0°C range until the end of the glaciation.

These results have several implications. The seawater $\delta^{18}\text{O}$ record obtained by combining the benthic foraminiferal records of the Norwegian Sea and of core V19-30 (Fig. 5) presents a smaller glacial/interglacial shift (1.1‰) than previous estimates^{12,30} which neglected the deep-water temperature changes in the Pacific and Indian Ocean. Overestimation of the glacial Norwegian Sea/Pacific Ocean gradient in water $\delta^{18}\text{O}$ could bring this figure to 1.3‰, but the presently admitted figure of 1.6‰ is incompatible with our data. Using a very different approach, the comparison between detailed records of sea level changes during the last climatic cycle on raised coral reefs and benthic foraminiferal $\delta^{18}\text{O}$, Chappell and Shackleton³¹ in the Huon Peninsula (New Guinea), and Dodges and co-authors³² in Haiti, have remarked that reconciliation of recorded high values of sea-level and Pacific $\delta^{18}\text{O}$ records implied that approximately 0.5‰ of the mean 1.6‰ glacial/interglacial foraminiferal $\delta^{18}\text{O}$ shift must be attributed to a deep-water temperature decrease during the last glacial. This result confirms the different hypotheses involved in our reconstruction.

This study has another implication. The 2°C cooling of the deep world ocean during the last glaciation has increased the solubility of CO_2 and O_2 . It drastically changes vertical density distribution and thermohaline circulation. But quantitative estimation of the resulting effects on the Earth's climate will demand a better knowledge of the three-dimensional ocean dynamic and air-sea exchanges. It is also by this method that it will be possible to check accurately the hypothesis involved in the present work, especially the temperature stability during high-latitude bottom-water formation.

This work was supported by Programme National d'Etude de la Dynamique du Climat, Commissariat à l'Energie Atomique, Mission Scientifique des Terres Australes et Antarctiques Française, Centre National de la Recherche Scientifique, and Ministère de l'Environnement. Sediment samples came from Centre des Faibles Radioactivités and Lamont-Doherty Geological Observatory core collections (the latter is supported by NSF). We acknowledge support from J. Antignac, H. Leclaire, and B. Le Coat for the isotope analyses, and E. Kaltnecker for drafting. This research has been helped by discussions and informations provided by numerous colleagues, and particularly N. J. Shackleton and R. Fairbanks. The manuscript has been improved following a review by D. Oppo.

Received 24 November 1986; accepted 16 April 1987.

- Emiliani, C. *J. Geol.* **63**, 538–578 (1955).
- Urey, H. C. *J. chem. Soc.* 562–581 (1947).
- Epstein, S., Buchsbaum, R., Lowenstam, H. & Urey, H. C. *Geol. Soc. Am. Bull.* **64**, 1315–1325 (1953).
- Shackleton, N. J. *Variations du Climat, au cours du Pleistocène* 203–209 (CNRS, Gif, 1974).
- Duplessy, J. C., Bé, A. W. H. & Blanc, P. L. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **33**, 9–46 (1981).
- Bouvier-Soumagnac, Y. & Duplessy, J. C. *J. foram. Res.* **15**, 302–320 (1985).
- Shackleton, N. J., Imbrie, J. & Hall, M. A. *Earth planet. Sci. Lett.* **65**, 233–244 (1983).
- Foster, T. D. & Middleton, J. H. *Deep Sea Res.* **27A**, 367–381 (1981).
- Lazier, J. R. N. *Deep Sea Res.* **20**, 341–353 (1973).
- Mantilla, A. W. & Reid, J. L. *Deep Sea Res.* **30**, 805–833 (1983).
- Ruddiman, W. F., Sancetta, C. D. & McIntire, A. *Phil. Trans. R. Soc. B* **280**, 119–142 (1977).
- Duplessy, J. C., Moyes, J. & Pujol, C. *Nature* **286**, 279–482 (1980).

13. Duplessy, J. C. & Shackleton, N. J. *Nature* **316**, 500-507 (1985).
14. Craig, H. & Gordon, L. in *Stable Isotopes in Oceanographic Studies and Palaeotemperatures* 9-131 (Consiglio Nazionale delle Ricerche, Pisa, 1965).
15. Kellogg, T. B. *Boreas* **9**, 115-117 (1980).
16. Streeter, S. S., Belanger, P. E., Kellogg, T. B. & Duplessy, J. C. *Quat. Res.* **18**, 72-90 (1982).
17. Duplessy, J. C. *et al. Quat. Res.* **21**, 225-243 (1984).
18. Crousset, F. & Duplessy, J. C. *Mar. Geol.* **52**, M-11-M17 (1983).
19. Duplessy, J. C., Chenouard, L. & Vila, F. *Science* **188**, 1208-1209 (1975).
20. Duplessy, J. C. & Labeyrie, L. D. Long and Short Term Variability of Climate Lecture Notes in Earth Sciences Series (eds Wanner, H. & Siegenthaler, U.) (Springer, Berlin, in the press).
21. Zahn, R., Winn, K. & Sarnthein, M. *Paleoceanography* **1**, 27-42 (1986).
22. Labeyrie, L. D. & Duplessy, J. C. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **50**, 217-240 (1985).
23. Imbrie, J. *et al.* in *Milankovitch and Climate* NATO ASI Series C126 (eds Berger, A. L., Imbrie, J., Hays, J., Kukla, G. & Saltzman, B.) 269-306 (Reidel, New York, 1984).
24. Paterne, M., Guichard, F., Labeyrie, J., Gillot, P. Y. & Duplessy, J. C. *Mar. Geol.* **72**, 259-285 (1986).
25. Duplessy, J. C. *et al. Nature* **320**, 350-352 (1986).
26. Bard, E., Arnold, M., Duprat, J., Moyes, J. & Duplessy, J. C. *Climat. Dyn.* **1**, 101-112 (1987).
27. Savitsky, A. & Golay, M. J. E. *Analyt. Chem.* **36**, 1627-1639 (1964).
28. Labeyrie, L. D. *et al. Nature* **322**, 701-706 (1986).
29. CIIMAP project members *Quat. Res.* **21**, 123-224 (1984).
30. Shackleton, N. J. *Nature* **215**, 15-17 (1966).
31. Chappell, J. M. & Shackleton, N. J. *Nature* **324**, 137-140 (1986).
32. Dodge, R. E., Fairbanks, R. G., Benninger, L. K. & Murras, F. *Science* **219**, 1423-1425 (1983).
33. Dietrich, G. *Atlas of the Hydrography of the Northern North Atlantic Ocean* (Conseil Int. pour l'Exploration de la Mer, Charlottenlund Slot, 1969).
34. Wyrtki, K. *Oceanographic Atlas of the International Indian Ocean Expedition* (NSF, Washington DC, 1971).
35. Fuglister, F. C. *Atlantic Ocean Atlas from the International Geophysical year of 1957-1958* (Woods Hole Oceanographic Institution, Woods Hole, 1969).
36. *GEOSECS Pacific Expedition Vol 4* (Craig, H., Broecker, W. S. & Spencer, D. (eds) IDOE NSF, Washington DC, 1981).
37. O'Neil, J. R., Clayton, R. N. & Mayeda, T. K. *J. chem. Phys.* **51**, 5547-5558 (1969).

Myristylation of picornavirus capsid protein VP4 and its structural significance

M. Chow*, J. F. E. Newman^{†§}, D. Filman[‡], J. M. Hogle[‡], D. J. Rowlands[†] & F. Brown^{†||}

* Department of Applied Biological Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

† Wellcome Biotechnology Ltd, Pirbright, Woking, Surrey GU24 0NQ, UK

‡ Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, California 92037, USA

We have obtained evidence that poliovirus and other picornavirus particles are specifically modified by having myristic acid covalently bound to a capsid protein. The electron density map of poliovirus confirms the position of the myristate molecule and defines its location in the virus particle. Analogies with other myristylated proteins suggest that the myristate moiety in picornaviruses may be involved in capsid assembly or in the entry of virus into cells.

POST-TRANSLATIONAL modifications of eukaryotic host and viral proteins are important for proper localization, stability and function. Several recent reports have described acylation of viral proteins with fatty acids¹⁻³, which in many cases is responsible for localization of the viral protein on the plasma membrane of infected cells^{4,5}. Although the picornaviruses are non-enveloped they seem to interact specifically with cellular membranes during entry into the cell and also during virus particle assembly⁶. We report here that myristic acid (*n*-tetradecanoic acid) is covalently linked to the capsid protein VP4 and to its precursors VP0 and P1 (ref. 7) in poliovirus, bovine enterovirus (BEV), foot and mouth disease virus (FMDV), encephalomyocarditis virus (EMCV), and human rhinovirus (HRV), thus demonstrating myristylation in all four genera of the picornavirus family. In two cases, poliovirus and BEV, myristylation is shown to occur by an amide linkage to the N-terminus of VP4. In poliovirus the myristic acid moiety is also identified in electron density maps derived from high-resolution X-ray crystallographic studies. In all known picornavirus sequences, the N-terminus of VP4 is consistent with a consensus N-terminal sequence previously identified in other myristylated proteins; N-terminal myristylation may be a general property of the picornavirus family and be essential in their structural organization or replication.

Chemical evidence for myristylation of VP4

The family Picornaviridae consists of four genera; enterovirus (such as poliovirus and BEV), cardiovirus (such as EMCV), rhinovirus (HRV) and aphthovirus (FMDV)⁸. These non-enveloped RNA viruses are composed of 60 copies of each of four virus proteins (called VP1-4), forming a spherical shell, ~30 nm in diameter, with icosahedral symmetry; and a single-

stranded RNA genome of ~8,000 nucleotides, of positive polarity, polyadenylated at the 3' end and covalently linked at the 5' end to a small protein, VPg.

To test for attached fatty acids, poliovirus (serotype 1, Mahoney strain) was isolated from infected HeLa cells grown in the presence of either ³H-myristic acid or ³H-palmitic acid (*n*-hexadecanoic acid), added at various times from one day before infection to four hours after infection. The radiolabel co-purified on sucrose gradients with infectious virus isolated from cells labelled with either precursor, but the virus was labelled to much lower specific activity when grown with ³H-palmitic acid. The ³H-myristate- or ³H-palmitate-labelled poliovirus preparations were hydrolysed with 6 M HCl and the liberated fatty acids analysed by thin layer chromatography in the presence of authentic fatty acid markers. Only myristic acid was recovered from both the ³H-myristate- and ³H-palmitate-labelled viruses. A similar result was obtained with BEV (Fig. 1). These results show that myristate is specifically incorporated and rule out the possibility that the incorporated label is a metabolic product of the added myristate.

To characterize the fatty-acid modification in the virus particle, the protein from ³H-myristate-labelled poliovirus was analysed by SDS-polyacrylamide gel electrophoresis (PAGE)⁹ (Fig. 2a). The radioactivity always specifically co-migrated with the VP4 capsid protein, indicating that the myristic acid was covalently bound to VP4. Such modification could account for the anomalous retention on C18 reverse-phase HPLC columns of VP4, which requires unusually high methanol or isopropanol concentrations for its elution (M.C., unpublished data). To test the possibility that this modification is a general property of the family, similar experiments were carried out with BEV, EMCV, FMDV and HRV. Preparations of poliovirus and HRV labelled with ³H-myristate were grown in HeLa cells and BEV, EMCV and FMDV in BHK cells. Parallel cultures were grown with ³⁵S-methionine label to provide reference markers for the capsid proteins. Analysis of the proteins from purified viruses by SDS-

§ Permanent address: Department of Microbiology, Rhodes University, Grahamstown 6140, South Africa.

|| To whom correspondence should be addressed.

Fig. 1 Identification of the protein-linked fatty acid on VP4. BEV labelled with either ^3H -myristic acid or ^3H -palmitic acid was purified on sucrose gradients, hydrolysed *in vacuo* with 6 M HCl at 110 °C for 18 h, then extracted twice with toluene. The toluene extracts were concentrated and separated by ascending chromatography in lanes on C18 reverse-phase TLC plates (Whatman) with acetic acid/acetonitrile (1:1). Parallel lanes contained ^3H -fatty acid markers. Radioactivity in 0.5-cm cuts across each lane was determined using a liquid scintillation counter.

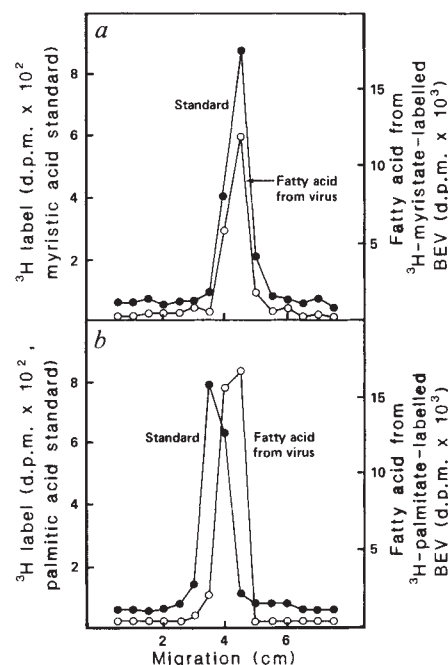
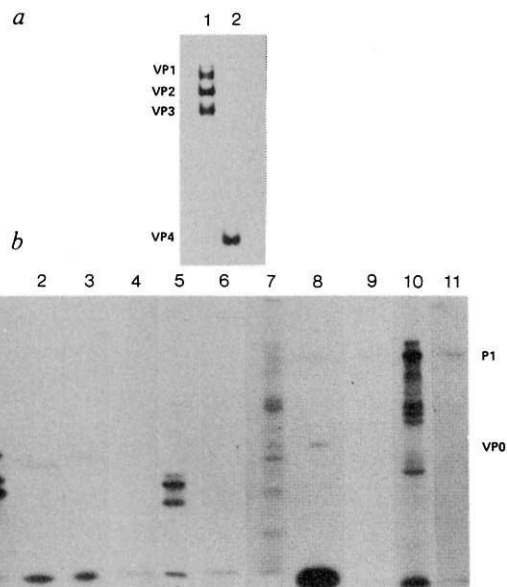


Fig. 2 SDS-PAGE of radiolabelled virus proteins. *a*, Poliovirus (^{35}S , lane 1; ^3H , lane 2). *b*, BEV (^{35}S , lane 1; ^3H , lane 2); EMC (^3H , lane 3); HRV (^3H , lane 4); FMDV (^{35}S , lane 5; ^3H , lane 6). FMDV infected cells (10^8) were incubated with 100 μCi ^{35}S -methionine or 1 mCi ^3H -myristic acid for 2 h, then lysed by heating at 100 °C for 2 min in disruption solution, (^{35}S , lane 7; ^3H , lane 8); ^3H -labelled myristylated virus induced proteins were precipitated with anti-VP1 antiserum for 1 h at 37 °C and then with Pansorbin (Calbiochem) for 4 h at 4 °C (lane 9). Rabbit reticulocyte lysate (100 μl), primed with FMDV-RNA was incubated for 35 min at 37 °C in the presence of 1 μCi ^{35}S -methionine or 10 μCi ^3H -myristic acid (^{35}S , lane 10; ^3H , lane 11). Note that equivalent proteins from the different viruses do not migrate at the same rate in SDS-PAGE.

Methods. Poliovirus and HRV infected HeLa cells (5×10^7 cells) and BEV, EMCV and FMDV infected BHK cells (10^8 cells) were labelled with 1 mCi ^3H -myristic acid or 100 μCi ^{35}S -methionine. Sucrose gradient purified viruses were disrupted by heating at 100 °C for 2 min in disruption solution (2% SDS and 2% β -mercaptoethanol) and separated by SDS-PAGE on 10% gels⁹.



PAGE demonstrated that VP4 capsid proteins of BEV (Fig. 2*b*, lane 2), EMCV (lane 3), HRV (lane 4) and FMDV (lane 6), were radiolabelled with ^3H -myristate. Myristylation of VP0, the precursor of VP2 and VP4, also occurred, although the amount of label was low because of the small number of copies of VP0 in virus particles. No radioactivity was associated with the other major capsid proteins. Thus myristate modification of VP4 is probably a general property of all picornaviruses.

During infection, the information encoded in the viral RNA is expressed as a single primary translation product which is the precursor for all viral proteins⁷. Capsid proteins are generated from this polypeptide by a series of specific cleavages. VP4 is initially found as the N-terminus of the capsid protein precursor P1, and after cleavage of P1 (to VP0, VP3 and VP1) as the N-terminus of VP0, the immediate precursor of VP4. That VP4 was covalently myristylated suggested that the precursors of VP4 might be similarly modified. Analysis of radiolabelled proteins from FMDV infected cells (Fig. 2*b*, lanes 7 and 8) and those produced *in vitro* in a cell-free lysate primed with FMDV RNA (Fig. 2*b*, lanes 10 and 11) demonstrated that P1 and VP0

were also myristylated. The identity of ^3H -myristylated P1 (which contains the capsid protein sequences) in an infected cell lysate was confirmed by its precipitation with anti-VP1 antiserum (Fig. 2*b*, lane 9). Similar results were obtained with cell lysates from poliovirus infected cells. The presence of myristic acid in the capsid protein precursors of FMDV, synthesized *in vivo* (Fig. 2*b*, lane 8) or *in vitro* (Fig. 2*b*, lane 11), indicates that this modification occurs as an early post-translational event.

A fatty acid molecule can be linked to a protein by an ester, thio-ester or amide bond. An amide bond can be distinguished from both ester and thio-ester linkages by its resistance to hydrolysis with 1 M hydroxylamine at either pH 7 or pH 10. The ^3H -myristate label could not be released from poliovirus VP4 by treatment with 1 M hydroxylamine at pH 7 or pH 10 as judged by the failure to elute it from SDS-PAGE slices. Moreover, the label could not be extracted with chloroform/methanol (2:1) from SDS-disrupted virus which had been treated with hydroxylamine. This stability indicates that the myristic acid moiety is linked to VP4 by an amide bond.

An amide linkage can be formed by myristylation of either

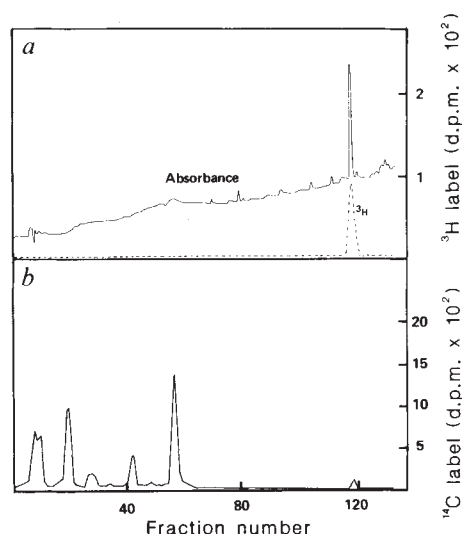


Fig. 3 Co-chromatography of *a*, ^3H -myristate and *b*, ^{14}C -labelled protein hydrolysate tryptic peptides from BEV VP4 with synthetic N-terminal peptide. Dried polyacrylamide gel slices containing labelled VP4 were digested with trypsin ($30\text{ }\mu\text{g}$ in 0.5 ml 50 mM ammonium bicarbonate $\text{pH } 8.0$ for 24 h at 37°C^{10}). Peptide digests were separated from gel fragments by filtration through a $0.45\text{-}\mu\text{m}$ Millipore membrane. In *a*, the digested ^3H -myristate labelled VP4 was mixed with $40\text{ }\mu\text{g}$ of synthetic N-terminal peptide (My-Gly-Ala-Gln-Leu-Ser-Arg) before injection onto a μ -Bondapak C18 column (Waters Associates). Elution was with a gradient of $10\text{--}60\%$ acetonitrile in 0.1% trifluoroacetic acid (TFA) and 0.1% morpholine over 60 min . Column effluent was monitored at 215 nm and 0.5-ml fractions collected for radioactivity determination.

the ϵ -amino group of a lysine residue in the protein or by myristylation of the free α -amino group of the N-terminal amino acid. In all myristylated proteins described previously the fatty acid has been shown to be linked by an amide bond to the N-terminus of the protein. Furthermore, in several picornaviruses, the N-terminus of VP4 is known to be blocked and inaccessible to Edman degradation. It therefore seemed likely that the myristate is linked to the α -amino group of VP4. This hypothesis was tested by analysis of tryptic peptides derived from ^3H -myristylated VP4. BEV was chosen for these experiments because its RNA sequence indicates that tryptic digestion would produce a myristylated hexapeptide because an arginine is predicted six residues from the N-terminus (S. J. Martin, personal communication).

BEV was radiolabelled with either ^3H -myristic acid or ^{14}C -labelled protein hydrolysate. The VP4 protein from each virus preparation was isolated by SDS-PAGE, digested with trypsin¹⁰ and the peptides separated by reverse-phase HPLC (Fig. 3). When isolated from virus grown in the presence of ^3H -myristate and digested with trypsin, VP4 generated a single labelled peptide that co-eluted with a synthetic peptide My-Gly-Ala-Gln-Leu-Ser-Arg (where My represents myristate), corresponding to the N-terminus of VP4 (S. J. Martin, personal communication). As expected from the amino-acid sequence for BEV (S. J. Martin, personal communication), VP4 labelled with ^{14}C -protein hydrolysate generated at least five peptides after digestion with trypsin, the most hydrophobic co-eluting with the synthetic N-terminal

peptide. The only amino acid in this sequence which can form an amide bond with myristic acid is the N-terminal glycine. These experiments provide further evidence that myristylation occurs on the α -amino group of the N-terminal glycine of VP4 in BEV.

Location of myristic acid in poliovirus

The structure of the Mahoney strain of type-1 poliovirus has been determined at $2.9\text{-}\text{\AA}$ resolution¹¹. In the initial $2.9\text{-}\text{\AA}$ electron density map the density at the N-terminus of VP4 was not sufficiently defined to allow an unambiguous identification of the N-terminal residues. Subsequent crystallographic refinement of the Mahoney structure (D.F. and J.M.H., unpublished observations) has resulted in a significant improvement in this area of the map. The refined map shows the side chains of the N-terminal residues clearly, allowing an unambiguous identification of the terminal glycine. Additional electron density extending from the α -amino group of the terminal glycine is entirely consistent with this residue being myristylated (Fig. 4). Specifically, there is partially resolved density for the carbonyl oxygen of the amide bond between C1 of the myristate and the terminal glycine, and the remaining density can be completely accounted for by 13 additional carbons with standard bond lengths and angles for saturated hydrocarbons and with torsion angles constrained to the preferred staggered conformations. At C12 there is a branch in the density indicating that the final two carbon atoms occupy two similarly populated positions corre-

Table 1 Amino-terminal sequences of myristylated proteins

Protein		Sequence							
Polio (Type 1) ¹⁹	VP4	Gly	Ala	Gln	Val	Ser	Ser	Gln	Lys
BEV (VG-27)*	VP4	Gly	Ala	Gln	Leu	Ser	Arg	Asn	Thr
EMCV ²⁰	VP4	Gly	Asn	Ser	Thr	Ser	Ser	Asp	Lys
FMDV (01-K) ²¹	VP4	Gly	Ala	Gly	Gln	Ser	Ser	Pro	Ala
HRV (1B)†	VP4	Gly	Ala	Gln	Val	Ser	Arg	Gln	Asn
RSV ⁴	60 ^{src}	Gly	Ser	Ser	Lys	Ser	Lys	Pro	Lys
MuLV ¹	15	Gly	Gln	Thr	Val	Thr	Thr	Pro	Leu
FeLV ^{3,22}	15	Gly	Gln	Thr	Ile	Thr	Thr	Pro	Leu
BaeV ^{3,23}	12	Gly	Gln	Thr	Leu	Thr	Thr	Pro	Leu
HTLV ^{3,24}	19	Gly	Gln	Ile	Phe	Ser	Arg	Ser	Ala
BLV ^{3,25}	15	Gly	Asn	Ser	Pro	Ser	Tyr	Asn	Pro
HIV ^{1,26,27}	17	Gly	Ala	Arg	Ala	Ser	Val	Leu	Ser
cAMP protein kinase ²⁸		Gly	Asn	Ala	Ala	Ala	Ala	Lys	—
Calcineurin B ²⁹		Gly	Asn	Glu	Ala	Ser	Tyr	Pro	Leu

* S. J. Martin; personal communication.

† G. Stanway; personal communication.

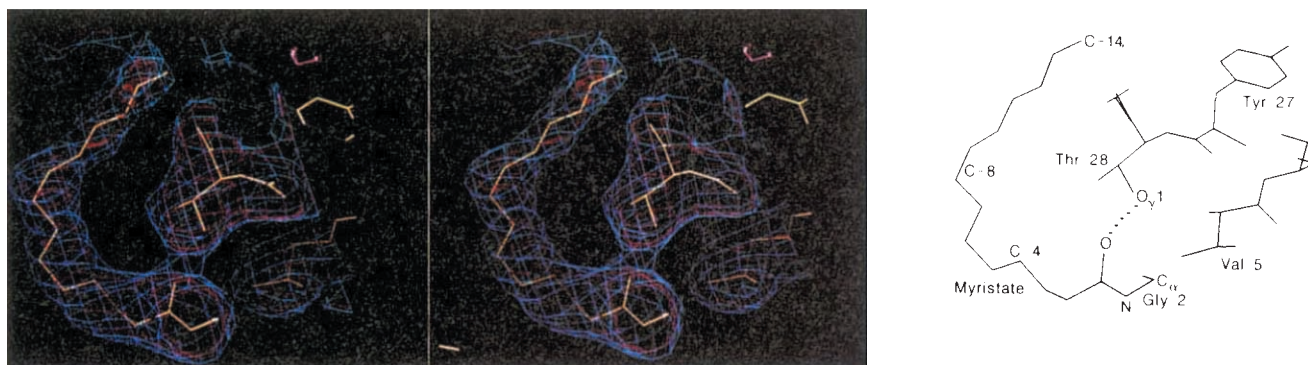


Fig. 4 A portion of the electron density map of the Mahoney strain of type 1 poliovirus showing myristate bound to the N-terminal glycine of VP4. The two panels on the left are a stereo pair with (right) a diagram of the same portion. The atomic model (yellow) has been constructed to match the electron density contours (blue and red). Note that Gly 2 and the attached myristate group are from one copy of VP4, whereas the other residues shown are from a fivefold symmetry-related copy of VP4. The N-terminal residue of VP4 is identified as Gly 2 to be consistent with codon identification in the genome sequence. The initiating methionine (Met 1) is removed from the precursor polypeptide in an early post-translational modification. The identification of the N-terminal glycine is based on the unambiguous correspondence between amino-acid sequence and electron density features for residues 2–16. (Residues 17–23 form a disordered loop that connects two well-ordered antiparallel β -strands.) It is clear from the electron density that there is a planar amide bond between the α -amino nitrogen of the terminal glycine of VP4 and C-1 of myristate. The myristyl moiety extends in an arc from the lower centre to the upper left. Note the partially resolved density for the carbonyl oxygen of the myristyl moiety. The dotted line in the diagram indicates the hydrogen bond between this carbonyl oxygen and the hydroxyl group of Thr 28.

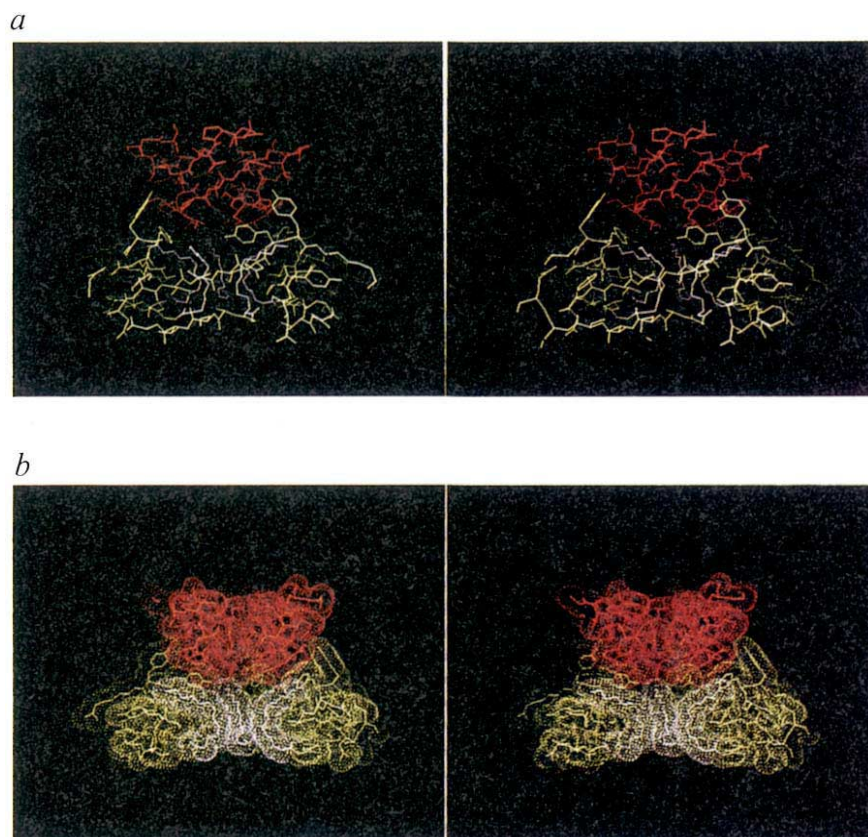


Fig. 5 Stereo views showing the interactions between five copies each of the N-terminus of VP3 and the myristylated N-terminus of VP4 in poliovirus. In these views, the fivefold symmetry axis of the poliovirus particle is vertical; the inside of the particle is downward; the outside of the particle is upward. *a*, Skeletal representation of the atomic model. At the bottom, shown in green, are five copies of the two-stranded antiparallel β -sheet at the N-terminus of VP4 (residues 4–8 and 25–29). This portion of VP4 lies roughly 100 Å from the particle centre and forms part of the inner surface of the protein shell. At the top, shown in red, five copies of the first 10 residues at the N-terminus of VP3 form a twisted tube of parallel β -structure. The top of the β -tube extends to about 130 Å from the particle centre. Near the fivefold axis, the outer surface has a maximum radius of 165 Å. The myristyl moiety (white) participates directly in the interaction between the N-termini of VP3 and VP4. *b*, Van der Waals surfaces calculated from the atomic model and sectioned parallel to the page show the extensive contacts of the myristyl hydrocarbon chains with both VP3 and VP4.

sponding to two of the three possible staggered conformations about the C12–13 bond. Recently a similar feature at the N-terminus of VP4 has been observed in electron density maps of the Sabin strain of type 3 poliovirus (D.F., M.C. and J.M.H., unpublished observations). These observations confirm that the N-terminal glycine of VP4 is linked by an amide bond to myristate and define for the first time the structure of a fatty-acid moiety in the three-dimensional structure of an acylated protein.

Myristate seems to form an integral part of the protein shell

of the virion. Around the fivefold axis, five copies of the N-terminus of VP3 intertwine to form a twisted tube of parallel β -structure. This compact structure is flanked on the inner surface of the shell by five (symmetry related) short pieces of two-stranded antiparallel β -sheet formed by residues near the N-termini of VP4. Figure 5 shows that as the five myristyl groups extend from the N-termini of VP4, they first come together in a tight cluster near the fivefold axis and then splay apart to cradle the twisted tube formed by the N-termini of VP3. Each

myristyl groups interacts with several amino-acid side chains from VP4 and VP3. Its contact with the Leu residue at position 2 of VP3 is particularly interesting because this hydrophobic residue (which is conserved in all poliovirus sequences and in most other picornaviruses) is protected from solvent by the hydrocarbon cluster.

Discussion

That myristate modification of VP4 is found in representative viruses from all four picornavirus genera shows that it is a highly conserved feature of the family and suggests that it may have a crucial function. N-terminal myristylation has also been described for several other proteins (Table 1). These include a number of retroviral proteins and cellular enzymes (see references in Table 1). The observed myristylation of enzymes, from presumably normal cells, indicates that N-terminal myristylation can be performed by cellular enzymes, but leaves open the possibility of contributions by viral enzymes or regulating factors. Amino-acid sequences necessary for myristylation of proteins are not known, but we suggest that the sequence: Gly-X-X-Ser/Thr forms the consensus sequence for myristylation, where X represents other amino acids. The N-terminal amino-acid sequences of VP4 from those viruses of each genus of the family Picornaviridae which we have examined and of other myristylated proteins are shown in Table 1. All the proteins have a serine or threonine at the fifth residue from the N-terminus with the exception of cAMP kinase. The alignment shown required a re-assessment of the proposed cleavage point between the leader protein and P1 of FMDV (ref. 12). Although there is little overall homology between these sequences, the conserved features are consistent with the proposed consensus sequence. The consensus sequence is present at the N-terminus of VP4 in all the picornaviruses for which information is available.

The ubiquity of the consensus sequence extends even to hepatitis A virus, a picornavirus, which despite a genome organization similar to other picornaviruses, differs in some of its properties. In particular there is conflicting evidence regarding the presence and size of VP4. Coulepis *et al.*¹³ reported finding a VP4 protein with a relative molecular mass (M_r) of 7,000–14,000 (7–14K) in purified virus particles, but sequence evidence suggests that VP4, if present, would only be 24 amino acids long and have M_r 2.5K (ref. 14). The amino-acid sequence data¹⁴ show that the proposed consensus sequence for myristylation (Gly-X-X-X-Ser/Thr), where X represents some amino acid, occurs seven amino acids downstream from the initiating codon of the polypeptide. If cleavage and myristylation occur in HAV it can be predicted that the VP4 of this virus would contain only 17 amino acids. This would be preceded by the precursor protein by a seven amino-acid leader sequence.

The proposed common N-terminal myristylation of VP4 among the picornaviruses is especially interesting because of differences in processing of the N-terminus. In the entero- and rhinoviruses the N-terminal glycine is produced by removal of

the initiating methionine, whereas in the cardio- and aphthoviruses (and perhaps HAV) the N-terminal glycine is produced by proteolytic removal of a leader peptide. Processing of a terminal methionine or leader peptide is also required in previously described myristylated proteins. It will therefore be interesting to determine whether myristylation is coupled with a particular type of cleavage.

The position of the myristate chain in the poliovirus structure suggests several possible functions. The myristate may be important for capsid assembly. The twisted tube formed by the five N-termini of VP3 is similar to the ' β -annuli' observed in the T = 3 plant viruses around which the assembly of the T = 3 capsid appears to nucleate^{15,16}. Similarly the formation of pentameric aggregates of VP1, VP3 and VP0 is an intermediate step in picornavirus capsid assembly and the β -tube is a structure that can form only on pentamer assembly and may be stabilized by myristate. The interaction with myristate may be important in stabilizing the twisted tube, promoting the conformational changes in the N-terminus of VP3 necessary for formation of the β -tube during capsid assembly. To initiate capsid assembly myristate modification may also be important for the localization of the capsid protein precursor P1 on the rough membranes. Myristylation of Pr65 of Moloney murine leukaemia virus has recently been shown to be essential for its localization on the plasma membrane and for the assembly of virus particles⁵.

The myristate could also be important in the early stages of infection, during binding and entry of virus particles into cells. Binding of picornaviruses and their subsequent elution from the cell-surface receptors always yields intact particles minus the VP4 protein¹⁷. Thus one of the earliest steps in the uncoating process, which is receptor- and membrane-dependent, is thought to be the specific relocation of VP4 in the capsid and its release in the case of abortive attachment. The modification of VP4 with myristate may facilitate this relocation and may direct VP4 to the cell or vesicle membrane where it could function in the subsequent passage of the virus across the membrane.

Finally, the sequestering of the myristyl moiety in the mature virion may control capsid-membrane interactions in space and time. Thus, during assembly the formation of the β -tube—myristyl-VP4 structure may serve to release mature pentamers, empty capsids, or virions from membranes. Similarly, the receptor-mediated externalization of VP4 from the virus particle may control the exposure of a membrane attachment or penetration signal. Further studies will be necessary to clarify the exact function of the myristate moiety.

Since this paper was submitted for publication, the occurrence of myristic acid covalently linked to VP2, a capsid protein of SV40 and polyoma virus, has been described¹⁸. These authors consider that with these viruses also the myristic acid modification may be involved in the assembly of virus particles.

We thank Mr. A. Syred for the peptides and HPLC analyses and Dr. G. Stanway for the then unpublished sequence of HRVIB. The work was supported in part by a grant from NIH and an NSF grant to M.C.

Received 17 February; accepted 30 April 1987.

- Henderson, L. E., Krutzsch, H. C. & Oroszlan, S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 339–343 (1983).
- Garber, E. A., Krueger, J. G., Hanafusa, H. & Goldberg, A. R. *Nature* **302**, 161–163 (1983).
- Schultz, A. & Oroszlan, S. *Virology* **133**, 431–437 (1984).
- Pellman, D., Garber, E. A., Cross, F. R. & Hanafusa, H. *Nature* **314**, 374–377 (1985).
- Rein, A., McClure, M. R., Rice, N. R., Luftig, R. B. & Schultz, A. M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7246–7250 (1986).
- Madhus, I. H., Olsnes, S. & Sandvig, K. *J. cell Biol.* **98**, 1194–1200 (1984).
- Rueckert, R. R. & Wimmer, E. *J. Virol.* **50**, 957–959 (1984).
- Matthews, R. E. F. *Intervirology* **17**, 1–199 (1982).
- Laemmli, U. K. *Nature* **227**, 680–685 (1970).
- Elder, J. H., Pickett, R. A., Hampton, J. & Lerner, R. A. *J. biol. Chem.* **252**, 6510–6515 (1977).
- Hogle, J. M., Chow, M. & Filman, D. J. *Science* **229**, 1358–1365 (1985).
- Clarke, B. E. *et al. J. gen. Virol.* **66**, 2615–2626 (1985).
- Coulepis, A. G., Locarnini, S. A., Westaway, E. G., Tannock, G. A. & Gust, I. D. *Intervirology* **18**, 107–127 (1982).
- Cohen, J. I., Ticehurst, J. R., Purcell, R. H., Buckler-White, A. & Baroudy, B. M. *J. Virol.* **61**, 50–59 (1987).
- Harrison, S. C. *Biophysics J.* **32**, 139–153 (1980).
- Sorger, P. K., Stockley, P. G. & Harrison, S. C. *J. molec. Biol.* **191**, 639–658 (1986).
- Crowell, R. L. & Landau, B. J. *Compreh. Virol.* **18**, 1–33 (1983).
- Streuli, C. H. & Griffin, B. E. *Nature* **326**, 619–622 (1987).
- Kitamura, N. *et al. Nature* **291**, 547–553 (1981).
- Palmenberg, A. A. *et al. Nucleic Acids Res.* **12**, 2969–2985 (1984).
- Beck, E., Forss, S., Strebl, K., Cattaneo, R. & Feil, G. *Nucleic Acids Res.* **11**, 7873–7885 (1983).
- Hampe, A., Laprevotte, I., Galibert, F., Fedele, L. A. & Sherr, C. J. *Cell* **30**, 775–785 (1982).
- Tamura, T. K. *J. Virol.* **47**, 137–145 (1983).
- Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3618–3622 (1983).
- Rice, N. R., Stephens, R. W., Burney, A. & Gilden, R. V. *Virology* **142**, 357–377 (1985).
- Oroszlan, S. in *Modern Approaches to New Vaccines* (eds Chanock, R. M., Ginsberg, H., Lerner, R. & Brown, F.) (Cold Spring Harbor Laboratory, New York, in the press).
- Ratner, L. *et al. Nature* **313**, 277–284 (1985).
- Carr, S. A., Biemann, A. K., Shoji, S., Parmelee, S. C. & Titani, K. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6128–6131 (1982).
- Aitken, A. *et al. FEBS Lett.* **150**, 314–318 (1982).

A superstring neutrino-dominated Universe?

D. V. Nanopoulos* & Keith A. Olive†

* Department of Physics, University of Wisconsin, Madison, Wisconsin 53706, USA

† School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA

Recent observations¹ indicate the existence of significant large-scale structure on scales $>40 h_0^{-1}$ Mpc. That this is also the typical scale for surviving structure in the hot dark matter models for galaxy formation (for example, neutrinos with masses $m_\nu \approx 30$ eV) may not be a coincidence (ref. 2; D. N. Schramm, Fermilab preprint 86/70, 1986). If neutrinos are to be reinstated as the leading candidates for the dark matter of the Universe, it becomes crucial to understand the origin of their mass in the context of the superstring³⁻⁹. Here we examine the possibility for small but non-zero neutrino masses in a class of superstring models⁶ giving rise to acceptable low-energy phenomenology.

Of the multitude of dark matter candidates, it is only the neutrino which is in fact a known and discovered particle. Indeed, the neutrino has played a crucial and central role in trying to understand galaxy formation and the missing mass problems. One of the most attractive features¹⁰ for massive neutrinos and cosmology is that for neutrino masses greater than a few eV, the Universe is dominated by neutrinos rather than nucleons so that an $\Omega = 1$ Universe is compatible with Big Bang nucleosynthesis¹¹ (Ω is the ratio of the total mass density to the critical mass density $\rho_c = 1.88 \times 10^{-29} h_0^2 \text{ g cm}^{-3}$ where $h_0 = H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ is the present value of the Hubble parameter). In addition, because of neutrino free streaming^{12,13}, the natural (minimal) scale for clustering (structure) is given by the Jeans length, $\lambda_J \approx 40 (m_\nu/30 \text{ eV}) \text{ Mpc}$ and hence the possibility for large-scale structure¹⁴⁻¹⁷, including filaments and voids.

It was recognized early, however, that neutrinos tend to produce too much large-scale structure¹⁸. Because the smallest nonlinear structures have length scales λ_J , and the correlation length (the scale at which the galaxy-galaxy correlation function is unity) is¹⁹ $5h_0^{-1}$ Mpc, galaxies must fragment out of the larger pancake-like objects. In such a model, galaxies form late^{17,20} (redshift $z < 1$) whereas quasars are seen as far out as redshifts $z \approx 3.8$. (For a comprehensive review see ref. 21).

Some of the difficulties with the hot dark matter model can be plausibly ameliorated^{22,23} (see also R. Cowsik and P. Ghosh, preprint, Tata Institute). It has also been suggested²⁴ that in a hybrid model of neutrinos and cosmic strings, structure on both large and small scales would be present. It has been shown²⁵, however, that cosmic strings are not compatible with any current model of inflation.

Recent interest in a neutrino-dominated universe is twofold. First there is the observation¹ of large-scale structure indicating that galaxies appear to lie on surfaces of spherical voids and second, there is a renewed interest²⁶ in neutrino masses as a solution to the solar neutrino problem. Whereas solutions based on vacuum oscillations of ν_e into other neutrino flavours required large mixing angles, the new solution²⁶ is based on resonant oscillations in matter²⁷ (the Sun). The solar neutrino problem can be resolved with small mixing angles provided $\Delta = m_{\nu_1}^2 - m_{\nu_2}^2 \approx 6 \times 10^{-5} \text{ eV}^2$. We stress, however, that the new observations do not require the revitalization of the neutrino model and that cold dark matter models may be consistent with these observations as well²⁸.

A common source for small neutrino masses is provided via the 'see-saw' mechanism²⁹ which, in the diagonalization of the

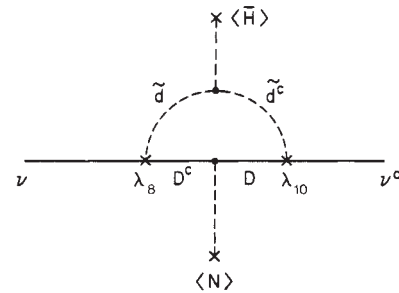


Fig. 1 Possible one-loop contribution to a Dirac neutrino mass.

neutrino mass matrix entails (in the simplest case) two eigenstates M , and m^2/M , where M is a large Majorana mass and m is a typical fermion Dirac mass. In superstring models, such a mechanism is, in general, not possible because of the absence of Majorana mass terms for neutrinos. Here we examine the possibility that neutrinos with small Dirac masses arising from one-loop radiative corrections in a minimal superstring model⁶ can satisfy the requirements necessary to resolve the solar neutrino problem while at the same time yielding an effectively stable ν_τ with a mass ~ 30 eV so that the Universe is neutrino dominated with $\Omega = 1$ and hence the possibility for large-scale structure in the Universe. We will also comment on the cosmological bounds on Dirac masses of this type.

If we accept the possibility that superstring theories are the all-encompassing theories of fundamental interactions, then we should in principle be able to calculate something as simple as a neutrino mass. If we take as our starting point³, a theory based on the gauge group $E_8 \times E_8$, we begin with a ten-dimensional space-time which must be compactified to four dimensions. The topology of the compactified dimensions is eventually what determines parameters such as neutrino masses in the low-energy limit. The problem is that we do not know the specific manifold of the compactified space. One generally assumes that this takes place on a Calabi-Yau manifold in order to preserve supersymmetry. During compactification one of the E_8 factors breaks³⁰ down to some subgroup of E_6 . Given constraints from low-energy phenomenology, one hopes that there exists a suitable manifold such that these constraints are satisfied.

In all these models, the matter fields belong to 27 representations of E_6 . That is, in addition to the 15 known fields: the quark doublet, $Q = (u, d)_L$; the conjugate up and down quarks, u^c, d^c ; the lepton doublet, $L = (\nu, e)_L$, and the conjugate electron e^c , there are two additional $SU(2)_L$ doublets: $H = (H^+, H^0)$, $\bar{H} = (\bar{H}^0, H^-)$, two colour triplets: D, D^c and two standard model singlets ν^c and N . In general, the superpotential contains eleven renormalizable couplings

$$\begin{aligned} f = & \lambda_1 \bar{H} L e^c + \lambda_2 \bar{H} Q d^c + \lambda_3 H Q u^c + \lambda_4 H \bar{H} N + \lambda_5 D D^c N \\ & + \lambda_6 D Q Q + \lambda_7 D^c u^c d^c + \lambda_8 D^c Q L \\ & + \lambda_9 D u^c e^c + \lambda_{10} D d^c \nu^c + \lambda_{11} H L \nu^c \end{aligned} \quad (1)$$

The first three terms in f are just the standard (supersymmetric) model Yukawa couplings and give rise to masses for the u, d quarks and the electron. (Our notation is generation suppressed so that u stands for u, c , and t and e stands for e, μ and τ and so on) through vacuum expectation values (VEV) for H and $\bar{H}$, $v = \langle 0|H|0 \rangle$ and $\bar{v} = \langle 0|\bar{H}|0 \rangle$. In addition, we expect a vacuum expectation value for the N field as well $x = \langle 0|N|0 \rangle$ which then gives masses to $H, \bar{H}$ and the D quarks through the λ_4 and λ_5 couplings. While in principle ν^c could also acquire a VEV there are phenomenological reasons³¹ to suspect that $\langle 0|\nu^c|0 \rangle = 0$. The remaining superpotential couplings ($\lambda_6 - \lambda_{11}$), if all present, are

problematic^{6,8}). For example, the combination of couplings λ_6 , λ_7 and λ_{10} would lead to fast proton decay. Neutrinos will receive unacceptably large Dirac masses unless $\lambda_{11} < 10^{-9}$. These problems can be avoided in a simple way by the inclusion⁶ of a Z_2 discrete symmetry (D, D^c, ν^c) $\rightarrow -1(D, D^c, \nu^c)$ causing λ_6 – λ_9 and λ_{11} all to vanish. In this case, one is left with three massless ν_L and ν_L^c states and an effectively stable proton.

For our purposes here, the above discrete symmetry is not very interesting since neutrinos remain massless. We still need λ_{11} to vanish because a vacuum expectation value for H yields a Dirac mass $\lambda_{11}v$ which is to be compared with the (u, c, t) masses given by λ_3v . One should also note that the superpotential in equation (1) cannot give rise to Majorana masses for neutrinos, and hence the see-saw mechanism is not feasible. (Majorana mass terms can be generated⁵) through non-renormalizable couplings but small masses also require the existence of an intermediate scale which is problematic for many reasons³²). Through a suitable choice of a discrete symmetry^{7,8}, or invoking topological reasons, it is possible to have only the couplings λ_6 , λ_7 and λ_{11} vanish while keeping λ_8 , λ_9 and λ_{10} non-zero. These couplings often yield Dirac masses through the one-loop diagram as shown in Fig. 1. Such a diagram leads to a mass^{7,8}

$$m_\nu = \frac{1}{16\pi^2} \lambda_8 \lambda_{10} m_d (m_S/m_D) \quad (2)$$

where m_d is the down quark mass matrix and $m_S \approx m_D \approx 1$ TeV are the scale of supersymmetry breaking and the D-quark mass ((N)) respectively.

If the couplings λ_8 and λ_{10} were generation independent, (although there is no reason to expect that they should be) $\lambda_8 \approx \lambda_{10} \approx 10^{-3}$ would yield masses⁷

$$m_{\nu_e} \approx 0.05 \text{ eV}, m_{\nu_\mu} \approx 1 \text{ eV} \text{ and } m_{\nu_\tau} \approx 30 \text{ eV} \quad (3)$$

making ν_τ an interesting candidate for the dark matter in the Universe. In addition neutrinos with these masses and couplings are relatively stable⁹,

$$\tau_\nu > 10^{24} \text{ s}.$$

It has been shown²⁶, however, that in order to use the resonant oscillation mechanism²⁷ to solve the solar neutrino problem one needs $\Delta \approx 6 \times 10^{-5} \text{ eV}^2$. For equal couplings this would require $\lambda \approx 10^{-4}$, which is the same order of magnitude as the Yukawa coupling of the s-quark. Equal couplings, however, would make $m_{\nu_e} < 1 \text{ eV}$ and hence cosmologically uninteresting. Of course we would normally expect the couplings λ_8 and λ_{10} to be generation dependent. Then for $\lambda_8 \approx \lambda_{10} \approx \lambda$ with $\lambda^{(1)} \approx 10^{-5}$, $\lambda^{(2)} \approx 10^{-4}$ and $\lambda^{(3)} \approx 10^{-3}$ where the superscripts refer to generations, we get

$$m_{\nu_e} \approx 3 \times 10^{-6} \text{ eV}, m_{\nu_\mu} \approx 10^{-2} \text{ eV} \text{ and } m_{\nu_\tau} \approx 30 \text{ eV} \quad (4)$$

giving us the possibility of solving both the solar neutrino problem and the dark matter problem (at least one of the dark matter problems).

We note at this point that typically these superstring models will contain an additional stable particle such as the photino³³. Depending on the details of the model, the photino (a cold dark matter particle) could also contribute substantially to the overall energy density of the Universe. This could lead to a hybrid model containing both hot and cold dark matter, and perhaps yield further benefits to the formation of structure on both large and small scales³⁴. We point out that this type of hybrid model is not fine-tuned. If a massive neutrino is present with a significant abundance, this would be in addition to the presence of the photinos which are present in large abundances for other reasons³³.

Because of the attractive features in this type of model and

the novel mechanism for generating small Dirac masses, it is worthwhile examining some of the consequences as well as cosmological constraints. One of the main consequences of such a model is that we would predict the existence of three light Dirac neutrino states. The quantum numbers of ν^c are such that it is a singlet under the low-energy gauge group (in fact it is also a $SU(5)$ singlet as well). Thus its gauge interactions must lie outside the standard model. Neutral current interactions would then be mediated by an extra Z_E gauge boson (there is always at least one extra present in superstring models³⁻⁶). Furthermore because of the absence of anomalous neutral current interactions and the excellent agreement between the mass of the Z gauge boson M_Z and the standard model prediction we know⁶ that the mass of the extra Z_E gauge boson $M_{Z_E} > 230 \text{ GeV}$ making the right-handed interactions weaker than the standard left-handed interactions. Because of this difference in the interaction strength, the standard cosmological limits on light neutrino masses are modified³⁵.

Ordinarily we would write for the mass density in neutrinos

$$\rho_\nu = \sum_\nu (g_\nu/2) m_\nu n_\nu = \sum_\nu (g_\nu/2) m_\nu (3/11) n_\gamma \quad (5)$$

where the sum is over all light ($m_\nu < \text{MeV}$) neutrino flavors, g_ν is the number of spin degrees of freedom ($g_\nu = 4$ for a Dirac state) and we have used the fact that the number density of neutrinos n_ν is related to the number density of blackbody photons by a factor of $(3/4) \times (4/11)$ (the factor of $3/4$ is due to the difference between Fermi and Bose statistics and the $4/11$ is due to the enhancement in the number of photons due to $e^\pm$ annihilation in the early Universe when $T \approx m_e$). Taking the temperature of the microwave background radiation $T_0 = 2.7 \text{ K}$ we can rewrite equation (5) as

$$\Omega_\nu h_0^2 = \sum_\nu (g_\nu/2) (m_\nu/100 \text{ eV}) \quad (6)$$

and the limit $\Omega_\nu h_0^2 < 1$ implies the upper limit on the sum of Dirac masses $\sum m_\nu < 50 \text{ eV}$.

Because the right-handed states interact more weakly than the left-handed states, they decouple from the thermal background earlier at a higher temperature^{36,37}, the abundance (number density) will be reduced by a factor $(T_\nu/T_{\nu_L})^3 \approx g(T_{d_L})/g(T_{d_R})$ where $g(T_{d_{L,R}})$ is the total number of relativistic degrees of freedom at the left (right) decoupling temperature. In addition T_{d_L} and T_{d_R} are related by $(T_{d_R}/T_{d_L})^3 \approx [g(T_{d_R})/g(T_{d_L})]^{1/2} (M_{Z_E}/M_Z)^2$ and hence the expression for Ω_ν becomes³⁵

$$\Omega_\nu h_0^2 = \sum_\nu (1 + (T_\nu/T_{\nu_L})^3) (m_\nu/100 \text{ eV}) \quad (7)$$

In the limit of extremely weak right-handed interactions we have $\sum m_\nu < 100 \text{ eV}$.

Besides the change in the cosmological limit on the neutrino mass, it has been pointed out^{37,38} that the existence of six light two-component neutrino states implies a lower limit to M_{Z_E} much stronger than that due to present-day accelerator limits. Big bang nucleosynthesis¹¹ establishes the limit $N_\nu < 4$ where N_ν is the number of two component neutrinos with full strength weak interactions. In our case, we have six neutrino states but three of them with weaker interactions. It suffices (refs 37, 38; D.V.N. and K.A.O., CERN preprint TH-4530/86, 1986) that $3(T_\nu/T_{\nu_L})^4 \leq 1$ or $(T_\nu/T_{\nu_L}) < 0.76$ and

$$M_{Z_E} > 500 \text{ GeV} \quad (8)$$

$$\Omega_\nu h_0^2 < \sum_\nu (m_\nu/70 \text{ eV}) \quad (9)$$

If we saturate the bound from nucleosynthesis and take $M_{Z_E} \approx 500 \text{ GeV}$ the superstring model we considered above implies that, from equation (4) $\sum m_\nu = 30 \text{ eV}$ and that $\Omega_\nu h_0^2 \approx 0.4$ which for $\Omega_\nu = 1$ corresponds to a present value of the Hubble para-

meter $h_0 \approx 0.65$ which is in remarkable agreement with the observed range of $0.4 < h_0 < 1$.

We have considered the possibility of generating small but cosmologically significant neutrino masses. If the recent evidence¹ for structure on large ($\lambda > 40$ Mpc) scales holds the test of time, it may indeed be necessary to revive the neutrino dominated Universe (ref. 2; D. N. Schramm, Fermilab preprint 86/70, 1986). We have shown that generating neutrino masses ≈ 30 eV providing the desired large-scale structure is possible in the context of superstring inspired models. In spite of the

difficulties in finding a specific compactification manifold space with the right properties so as to provide the correct discrete symmetry and hence the coupling necessary to make $m_\nu \neq 0$, we can at the same time solve the solar neutrino problem, generate large-scale structure, satisfy $\Omega = 1$, all with an acceptable value of $h_0 = 0.65$.

The work of D.V.N. was supported in part by the DOE and in part by the University of Wisconsin Research Committee with funds granted by the Wisconsin Alumni Research Foundation. The work of K.A.O. was supported in part by the DOE.

Received 1 December 1986; accepted 18 March 1987.

1. de Lapparent, V., Geller, J. F. & Huchra, J. P. **302**, L1-L5 (1986).
2. Melott, A. L. *Astrophys. J.* (in the press).
3. Candelas, P., Horowitz, G. T., Strominger, A. & Witten, E. *Nucl. Phys. B* **258**, 46-74 (1985).
4. Witten, E. *Nucl. Phys. B* **258**, 75-100 (1985).
5. Derendinger, J. P., Ibanez, L. & Nilles, H. P. *Nucl. Phys. B* **267**, 365-414 (1985).
6. Ellis, J., Enqvist, K., Nanopoulos, D. V. & Zwirner, F. *Nucl. Phys. B* **276**, 14-70 (1986); *Mod. Phys. Lett. A* **1**, 57-69 (1986).
7. Masiero, A., Nanopoulos, D. V. & Sanda, A. I. *Phys. Rev. Lett.* **57**, 663-666 (1986).
8. Campbell, B., Ellis, J., Enqvist, K., Gaillard, M. K. & Nanopoulos, D. V. *Nucl. Phys. B* (in the press).
9. Enqvist, K. & Maalampi, J. *Phys. Lett.* **180B**, 347-352 (1986).
10. Schramm, D. N. & Steigman, G. *Astrophys. J.* **243**, 1-7 (1981).
11. Yang, J., Turner, M. S., Steigman, G., Schramm, D. N. & Olive, K. A. *Astrophys. J.* **281**, 493-511 (1984).
12. Bond, J. R., Efstathiou, G. & Silk, J. *Phys. Rev. Lett.* **45**, 1980-1984 (1980).
13. Zeldovich, Ya. B. & Sunyaev, R. A. *Pisma Astr. Zh.* **6**, 457-464 (1980).
14. Peebles, P. J. E. *Astrophys. J.* **258**, 415-424 (1982).
15. Melott, A. *Man. Not. R. astr. Soc.* **202**, 595-604 (1983).
16. Bond, J. R. & Szalay, A. S. *Astrophys. J.* **274**, 443-468 (1983).
17. Frenk, C. S., White, S. D. M. & Davis, M. *Astrophys. J.* **271**, 417-430 (1983).
18. White, S. D. M., Frenk, C. S. & Davis, M. *Astrophys. J.* **274**, L1-L5 (1983).
19. Davis, M. & Peebles, P. J. E. *Astrophys. J.* **267**, 564-482 (1983).

20. Bond, J. R., Centrella, J., Szalay, A. S. & Wilson, J. in *Formation and Evolution of Galaxies and Large Structures in the Universe* (eds Audouse, J. & Tran Thanh Van, J.) 87-99 (Reidel, Dordrecht, 1983).
21. Primack, J. in *Proc. Int. School of Physics Enrico Fermi Varenna, Italy* (1984).
22. Schaeffer, R. & Silk, J. *Astrophys. J.* **281**, L13-L15 (1984).
23. Melott, A. L. *Astrophys. J.* **289**, 2-4 (1985).
24. Schramm, D. N. & Vittorio, N. *Comments Nucl. Particle Phys.* **15**, 1-8 (1985).
25. Vishniac, E. T., Olive, K. A. & Seckel, D. *Nucl. Phys. B* (in the press).
26. Mikheyev, S. P. & Smirnov, A. Yu. in *Proc. Xth Int. Workshop on Weak Interactions Finland* (1985).
27. Wolfenstein, L. *Phys. Rev. D* **17**, 2369-2374 (1978).
28. White, S. D. M. et al. *Astrophys. J.* **313**, 505-516 (1987).
29. Gell-Mann, M., Ramond, P. & Slansky, R. in *Supergravity* (eds Van P. Nieuwenhuizen & Freedman, D.) 315-321 (North Holland, Amsterdam, 1979).
30. Hosotani, Y. *Phys. Lett.* **126B**, 309-313 (1983).
31. Campbell, B., Ellis, J., Gaillard, M. K., Nanopoulos, D. V. & Olive, K. A. *Phys. Lett.* **180B**, 77-82 (1986).
32. Ellis, J., Enqvist, K., Nanopoulos, D. V. & Olive, K. A. *Phys. Lett. B* (in the press).
33. Campbell, B. et al. *Phys. Lett.* **173B**, 270-279 (1986).
34. Umemura, M. & Ikeuchi, S. *Astrophys. J.* **299**, 583-592 (1985).
35. Olive, K. A. & Turner, K. A. *Phys. Rev. D* **25**, 213-316 (1982).
36. Olive, K. A., Schramm, D. N. & Steigman, G. *Nucl. Phys. B* **180**, 497-515 (1981).
37. Ellis, J., Enqvist, K., Nanopoulos, D. V. & Sarkar, S. *Phys. Lett.* **167B**, 457-463 (1986).
38. Steigman, G., Olive, K. A., Schramm, D. N. & Turner, M. S. *Phys. Lett.* **176B**, 33-38 (1986).

Giotto measurements of cometary and solar wind plasma at the Comet Halley bow shock

A. J. Coates*, R. P. Lin†, B. Wilken‡, E. Amata§, K. A. Anderson†, H. Borg||, D. A. Bryant¶, C. W. Carlson†, D. W. Curtis†, V. Formisano§, K. Jockers‡, A. D. Johnstone*, A. Korth‡, D. A. Mendis#, H. Reme**, A. K. Richter‡, H. Rosenbauer‡, J. A. Sauvaud**, W. Stüdemann‡, M. F. Thomsen††, C. d'Uston** & J. D. Winningham‡‡

* Mullard Space Science Laboratory, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK

† Space Sciences Laboratory, University of California, Berkeley, California 94720, USA

‡ Max-Planck-Institut für Aeronomie, D-3411 Katlenburg-Lindau, FRG

§ Istituto di Fisica dello Spazio Interplanetario, 00044 Frascati, Italy

|| Kiruna Geophysical Institute, Umea University, 90187 Umea, Sweden

¶ Rutherford Appleton Laboratory, Chilton, Didcot, Oxfordshire OX11 0QX, UK

Department of Electrical Engineering and Computer Sciences, University of California, San Diego, California 92093, USA

** Centre d'Etude Spatiale des Rayonnements, CNRS-Toulouse University, 31029 Toulouse Cedex, France

†† Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

‡‡ Southwest Research Institute, San Antonio, Texas 78284, USA

The interaction of comets with the solar wind depends on the ionization of the heavy cometary neutrals (mostly H₂O and its dissociation products O, OH) which flow out from the nucleus, and the coupling of these newly produced cometary ions with the solar wind through its embedded magnetic field. The 'pick-up' of

these heavy ions slows the solar wind such that a shock may form. As expected, the structure of such a shock transition is highly complex because the gyroradius of a heavy cometary ion is much larger than that of a solar wind proton. Here we present a comparative study of the solar wind electrons and protons and the cometary pick-up ions measured by Giotto at the inbound crossing of the bow shock at comet Halley. We find a highly structured shock transition starting with a cometary ion 'foot' seen at a distance on the order of the cometary ion gyroradius upstream from a sharp decrease in the solar wind proton speed. The total solar wind thermal and magnetic pressure is dominated by the relatively small population of cometary ions throughout the shock region.

Initial measurements from the Giotto plasma instruments at the comet Halley bow shock have already been presented¹⁻⁴ and some analysis has been done on the individual data sets⁵⁻⁸. Here we present the first comparative study at the bow shock using results from the Johnstone plasma analyser (JPA) and the Reme plasma analyser (RPA) instruments^{9,10}. JPA consists of two sensors, the fast ion sensor (FIS) and the implanted ion sensor (IIS). The RPA results here are from the electron electrostatic analyser (EESA).

The FIS, designed to measure the solar wind ions with high time resolution, is a 270° electrostatic E/q analyser¹¹ with energy range 10 eV–20 keV q^{-1} . The sensor works in two modes, the 3D mode and the solar wind mode, synchronized to the spacecraft spin. Only data from the solar wind mode are presented here. This mode has a 45° × 45° field of view centred on the solar direction. The spacecraft spin axis is approximately in the ecliptic plane. Each solar wind measurement takes 500 ms and is repeated every two spacecraft spins or 8 s. The solar wind energy range covers one-quarter of the total energy range in 30 steps; the upper energy of this reduced range is controlled on board by an autoranging facility which tracks the peak energy of the solar wind. Under normal conditions the solar wind protons and α -particles are covered without gaps in velocity space. Bulk parameters are calculated on the ground for both the proton and the α -particle populations. But during the time of the bow-shock crossing, the proton peak is deflected, sometimes to

the edge of the field of view of this mode, which causes the autoranging to cycle its start energy through eight possible values until it finds the peak again. Density and temperature parameters are not trustworthy during the times when such deflection occurs⁶, but the velocity is trustworthy except during the autoranging cycling. Consequently only velocity is shown here.

The IIS, designed to measure cometary ions with mass discrimination, consists of five detector systems, each comprising an electrostatic E/q analyser with a $10^\circ \times 6^\circ$ field of view followed by a time-of-flight m/q analyser. The five detector systems are arranged to sample in the spacecraft polar angle, the spacecraft spin provides the azimuthal angle. The energy range is 88 eV to 88 keV covered in 32 steps with $\Delta E/E = 10\%$ and the time resolution for full energy coverage is 128 s. Data are transmitted for five mass groups with an azimuthal resolution of 22.5° . Here we use data from the third mass group, covering the range 6.5–20 AMU/ q , mainly consisting of ions from the water group. A lower energy cutoff of 1.2 keV is used to avoid contributions from solar wind protons^{3,12}. The gaps in angular and energy coverage are removed by increasing each number of measured counts by a factor that fills the bin in energy-angle space evenly at the measured count rate. Density is then calculated by integrating the cometary ion flux over all angles. No account is yet taken of the extension of the expected shell-like distribution in velocity space to the low energies not covered by the instrument. This may increase the values shown here by a factor of two or more in absolute terms; as the velocity space shells are almost uniformly filled at this time⁷, the relative fluctuations shown here are accurate.

The RPA1-EESA electron spectrometer measures the three-dimensional distribution of electrons between 10 eV and 30 keV¹⁰. This instrument consists of a novel symmetric quadri-spherical electrostatic analyser with a planar $360^\circ \times 4^\circ$ field of view¹³ followed by a micro-channel plate detector which divides the incoming particles into 17 angular sectors. A full, 4π steradian, three-dimensional distribution with $22.5^\circ \times 22.5^\circ$ resolution is obtained every half spin (2 s) of the spacecraft without gaps in angular coverage. These data are compressed on-board into a pitch angle distribution which is telemetered every 2 s.

The trajectory of the Giotto spacecraft makes an angle of 107° to the Sun-comet line and intersects the bow shock on the flank (see Fig. 1) at a distance along the spacecraft track of 1.16×10^6 km. The shock surface is assumed² to be a paraboloid with a flaring factor of two, giving a subsolar stand-off distance of 4×10^5 km. The spacecraft-nucleus relative speed was 68.4 km s^{-1} and the closest approach was 605 km on the sunward side of the nucleus.

The results from the three sensors are shown in Fig. 2. A 1-h period between 19:00 and 20:00 spacecraft event time (SCET) has been chosen for this study. The three panels show, from top to bottom, electron density, solar wind proton speed and cometary ion density respectively. The $>10 \text{ eV}$ electron density (upper panel of Fig. 2) is relatively constant at $\approx 5\text{--}6 \text{ cm}^{-3}$ before 19:20 SCET. The electron distributions appear typical of the solar wind, with a near-maxwellian core component below $\sim 50 \text{ eV}$. The unmeasured electrons below 10 eV would contribute an additional 20–30% to the measured density if the maxwellian is extrapolated to zero energy.

Small density fluctuations are observed between 19:00 and $\sim 19:20$ SCET, but they are of order $\Delta n/n \approx 10\%$, in contrast to the density fluctuations $\Delta n/n \approx 50\%$ or more observed by the ICE spacecraft upstream of comet Giacobini-Zinner's bow wave¹⁴. The lower level of fluctuations at comet Halley allows the bow shock and its structure to be identified, while at Giacobini-Zinner no clear shock signature was apparent in the electron plasma measurements although some shock-like features were seen¹⁵. A recent coordinated analysis of all the ICE data has confirmed the existence of a weak shock¹⁶. The

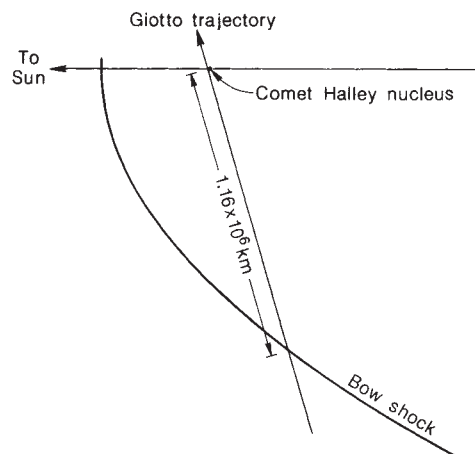


Fig. 1 Schematic view of the geometry as Giotto crossed the comet Halley bow shock.

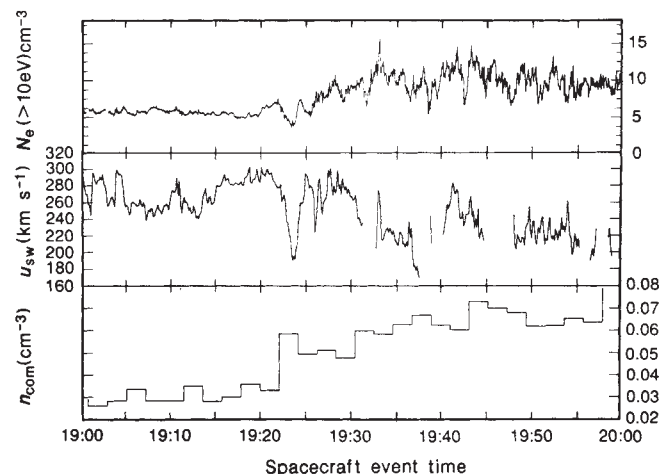
lower level of density fluctuations may be due to the lower rate of production of new pick-up ions at the comet Halley bow shock located at a cometocentric distance of $\sim 10^6$ km, in contrast to comet Giacobini-Zinner, where the bow wave is only $\sim 10^5$ km away from the nucleus. This point is discussed more fully elsewhere¹⁷.

The shock transition at Halley (Fig. 2) begins with a sharp rise in density from 5 to 7 cm^{-3} , starting at 19:20 SCET, reaching a broad peak at 19:21:30, followed by a dip to 4 cm^{-3} , a lower density than in the upstream solar wind, at $\sim 19:23$. A second rapid rise to $\sim 7 \text{ cm}^{-3}$ occurs at 19:24 SCET, followed by a drop to upstream values between 19:24:30 and 19:25:30, after which the density rises and remains elevated above upstream densities, although rapid fluctuations ($\sim$ few seconds timescale) are present. The same general behaviour was observed in the magnetic field strength¹, which increased from $\sim 5 \text{ nT}$ before 19:20 SCET to $\sim 8\text{--}9 \text{ nT}$ by 19:21:30, dropped to 4 nT at $\sim 19:23$ and then rose again at $\sim 19:24$ SCET to $\sim 7 \text{ nT}$, followed by a general increase with some rapid fluctuations.

The solar wind proton speed is shown in the central panel. Just upstream from the foreshock boundary (18:25 SCET) the speed was about 300 km s^{-1} (ref. 6); in the foreshock (18:25–19:15 SCET) variations in speed are evident between 245 and 300 km s^{-1} . At 19:15 SCET, the fluctuations cease and the average speed returns to just less than its value before the foreshock boundary. At 19:21 a decrease in speed occurs which lasts a total of 4 minutes; the most rapid decrease in speed starts at 19:22. The minimum speed in this feature is 190 km s^{-1} at 19:23:30. This dip in the solar wind ion velocity approximately coincides with the dip in both the electron density and the magnetic field strength to values below those in the upstream solar wind. At 19:25 the average speed has almost completely recovered but there are clear indications of waves of decreasing amplitude lasting until 19:30:30. At this time there is a sharp decrease in speed from 265 to 235 km s^{-1} over only one measurement (8 s), following which, the autoranging cycling mentioned earlier starts as the distribution leaves the field of view of the solar wind mode (visible as gaps on this plot). After this time the speed stays at an average level of about 225 km s^{-1} , except for a spike at $\sim 19:33$ and the interval $\sim 19:41$ to $\sim 19:44$ when the speed briefly recovers to a maximum of $\sim 280 \text{ km s}^{-1}$.

The lower panel shows the water group ion density. This quantity showed a gradual rise throughout the afternoon of 13 March with a more abrupt increase at the foreshock boundary¹⁸. The plot in Fig. 2 shows a large increase starting at 19:21:55 SCET (this time refers to the beginning of the 128 s cycle in which the IIS makes its measurement) with the average density showing a marked discontinuity at this time. Analysis of the

Fig. 2 Electron density (top panel), solar wind proton speed (middle panel) and cometary water-group ion density (bottom panel) for the hour on 13 March 1986 containing the comet Halley bow shock crossing.



individual cometary ion distributions at this time⁷ shows that the distributions clearly broaden and intensify here. The preliminary impression is that they are shell-like in velocity space in the vicinity of the shock both upstream and downstream, but the centre of the shell decreases in speed with the local solar wind speed. There is a noticeable peak in the instantaneous value of density for the one measurement starting at 19:21:55. The pick-up ion density appears to increase with each decrease in the solar-wind speed, for example at ~19:05, 19:12, 19:22, 19:31, 19:37 and 19:43 SCET. Note that apart from those increases the pick-up ion density is generally proportional to, and about 0.5% of, the solar wind electron density.

The shock transition thus begins with a region of solar-wind electron and magnetic field compression (19:20–19:22:30) followed by a region of rarefaction (19:22:30–19:24). The initial dip to ~190 km s⁻¹ in solar wind speed occurs in the rarefaction region. At the same time the pick-up ion density reaches a peak. Although the solar wind speed then recovers to its pre-shock level of ~300 km s⁻¹, after 19:25:30 the density of both solar wind electrons and pick-up ions is enhanced above pre-shock levels. The 'permanent' drop in solar wind speed does not occur until ~19:31 SCET, ~11 min and ~45,000 km after the first compression in the solar wind. This distance is more than five oxygen gyroradii ($r_{O^+} \sim 8,000$ km in the 6 nT upstream magnetic field) thick. The dip in the solar-wind speed at 19:23 is $\approx r_{O^+}$ wide. In contrast, the width of the permanent solar wind speed change at ~19:31 is at the limit of the FIS solar wind mode time resolution of 8 s, which corresponds to 550 km, or less than two gyroradii for solar-wind protons.

Referring to Fig. 1, it can be seen that for near-nominal angles (~45° to the Sun-comet line) of the interplanetary Parker spiral magnetic field and a nominal smooth shock shape, the shock will be quasi-perpendicular. The distances quoted above may then correspond to smaller numbers of oxygen gyroradii in the direction normal to the shock. Detailed analysis of the shock normal will be required to determine the scale size of the structure. It is clear, however, that the bow shock transition is highly structured and extends over perhaps several pick-up ion gyroradii but with sharp jumps in solar wind velocity on scales of a solar wind proton gyroradius. It is interesting to note that many of the features of this shock transition are observed in hybrid (fluid electron and particle ion) simulations of the process of heavy ion pick-up by the solar wind¹⁹, including the wave-like compressions and rarefactions and the rapid dips in solar wind speed followed by a return to unperturbed levels.

We now examine the energy density of the various plasma components and the magnetic field. Upstream of the shock (before 19:20 SCET) the electron energy density is $n_e kT_e \approx 200$ eV cm⁻³. The solar wind proton energy density in the solar-wind rest frame is $n_p kT_p \approx 50$ eV cm⁻³. For comparison the mag-

netic field energy density is 90 eV cm⁻³, making the ratio of thermal to magnetic energy density, β , for the solar wind ~3. Upstream of the shock the pick-up ions, assuming they are O⁺, have an energy density in the solar wind rest frame of ≈ 250 eV cm⁻³ even though their number density is only ≈ 0.03 cm⁻³. The total thermal and magnetic energy density is thus ~ 600 eV cm⁻³. The total plasma $\beta = 6$. Because of the uncertainty in the cometary ion density mentioned earlier β is likely to be even higher than this.

Downstream of the shock transition, after 19:31 SCET the energy density is ~ 400 eV cm⁻³ for the electrons, ~ 110 eV cm⁻³ for the protons, ~ 500 eV cm⁻³ for the pick-up ions and ~ 300 eV cm⁻³ for the magnetic field, giving a total thermal and magnetic energy density of at least $\sim 1,300$ eV cm⁻³. The plasma $\beta = 4$ and the jump in thermal and magnetic pressure across the shock is ~ 700 eV cm⁻³.

Thus about half of the plasma thermal pressure both upstream and downstream of the shock resides in the pick-up water-group ions. In particular, during the low-speed dip at 19:23 when the pressure contributions of all three solar-wind components (protons, electrons, magnetic field) decrease, the total appears to be balanced by the instantaneous increase in pick-up ion pressure. It is interesting to note that this cometary shock, where a minor species in terms of number density plays the dominant role in the pressure balance, is similar to supernovae shocks which presumably accelerate cosmic rays to high energies. The accelerated cosmic rays, although very low in density, are expected to exert a significant dynamic pressure in the shock^{20,21}.

R.P.L. acknowledges useful discussions with N. Omid. The work on the RPA data was supported by CNES by NASA and by the Bundesministerium für Forschung und Technologie. The work on the JPA data was supported by the SERC, the Bundesministerium für Forschung und Technologie, NASA, the US Department of Energy, the Swedish Board for Space Activities and the Piano Spaziale Nazionale of the Consiglio Nazionale delle Ricerche.

Received 26 January; accepted 18 March 1987.

1. Neubauer, F. M. *et al.* *Nature* **321**, 352–355, 1986.
2. Reme, H. *et al.* *Nature* **321**, 349–352, 1986.
3. Johnstone, A. *et al.* *Nature* **321**, 344–347, 1986.
4. Balsiger, H. *et al.* *Nature* **321**, 330–334, 1986.
5. Formisano, V. *et al.* *Adv. Space Res.* **6**, 229–233, 1986.
6. Coates, A. J. *et al.* *The Exploration of Halley's comet* Vol. 1 (eds Batrick, B., Rolfe, E. J. & Reinhard, R.) 263–268 (ESASP-250, Paris, 1986).
7. Wilken, B. *et al.* *The Exploration of Halley's comet* Vol. 1 (eds Batrick, B., Rolfe, E. J. & Reinhard, R.) 305–308 (ESASP-250, Paris, 1986).
8. Anderson, K. A. *et al.* *The Exploration of Halley's comet* Vol. 1 (eds Batrick, B., Rolfe, E. J. & Reinhard, R.) 259–261 (ESASP-250, Paris, 1986).
9. Johnstone, A. D. *et al.* *The Giotto mission—its Scientific Investigations* (eds Reinhard, R. & Batrick, B.) 15–32 (ESASP-1077, Noordwijk, 1986).
10. Reme, H. *et al.* *The Giotto mission—its Scientific Investigations* (eds Reinhard, R. & Batrick, B.) 33–52 (ESASP-1077, Noordwijk, 1986).

11. Johnstone, A. D. *et al.* *IEEE Trans. Nucl. Sci.* NS-32, 139–144, 1985.
12. Wilken, B. *et al.* *J. Phys. E* (submitted).
13. Carlson, C. W. *et al.* *Adv. Space Res.* 2, 67–70 (1982).
14. Bame, S. J. *et al.* *Science* 232, 356–361 (1986).
15. Hynds, R. J. *et al.* *Science* 232, 361–365 (1986).
16. Smith, E. J. *et al.* *The Exploration of Halley's comet* Vol. 3 (eds Battrock, B., Rolfe, E. J. & Reinhard, R.) 461–465 (ESASP-250, Paris, 1986).
17. Anderson, K. A. *et al.* *Geophys. Res. Lett.* (submitted).
18. Wilken, B. *et al.* *Astr. Astrophys.* (submitted).
19. Omid, N. & Winske, D. *Eos* 67, 1144 (1986).
20. Axford, W. I. *Proc. 17th Int. Cosmic Ray Conf* 12, 155–203 (Service de Doc. du CEN, Saclay, 1981).
21. Drury, L.O'C. *Rep. Prog. Phys.* 46, 973–1027 (1983).

Io plasma torus and the source of jovian narrow-band kilometric radiation

Dyfrig Jones

Space Plasma Physics Group, British Antarctic Survey, NERC,
Madingley Road, Cambridge CB3 0ET, UK

Narrow-band jovian kilometric radiation (nKOM) is distinguishable from broad-band jovian kilometric radiation (bKOM), both discovered by the Voyager spacecraft^{1–3}, by its generally lower intensity, its more restricted range of centre frequency, its narrower bandwidth, and its temporally less sporadic nature⁴. When observed at northern and southern magnetic latitudes nKOM polarization was predominantly left-handed and right-handed respectively, opposite to bKOM^{4,5}. At nearly the same longitude on consecutive jovian rotations nKOM sometimes lagged the magnetic field rotation⁴ by 3–5%; no such lag was detected for bKOM^{6,7}. It was proposed that bKOM is produced within $\sim 2^\circ$ of the magnetic equator in the outer Io plasma torus by conversion of electrostatic upper hybrid (ESUH) waves to ordinary (O) mode electromagnetic radiation via a radio window^{8,9}. Here we show that nKOM seems to be produced by the same conversion process, also operating in the Io torus, but at latitudes $\geq 8^\circ$. The difference in latitude for nKOM and bKOM sources stems from the condition required for mode conversion that the plasma density gradient should be normal to the magnetic field vector. The difference in these locations allows an explanation of most of the similarities and differences between the two radiations.

Voyager 1 (V1) approached Jupiter at 10:30 local time (LT) at jovicentric latitude 3.2° . Jupiter's magnetic axis is tilted 9.6° from its rotation axis, the north magnetic pole being at 202° central meridian longitude (CML). Therefore when V1 approached Jupiter, its magnetic latitude varied sinusoidally between 12.8° and -6.4° at Jupiter's rotation period of ~ 10 h.

The first observations of nKOM were made when the Voyager spacecraft were within 0.7 AU of their encounter with Jupiter^{1,5}. It is generally confined to 60–150 kHz with bandwidths typically 40–80 kHz, although some events are seen at 20 kHz and 175 kHz and with bandwidths < 20 kHz⁴. Figure 1 shows wave spectrograms composed from signals recorded on V1 by the Planetary Radio Astronomy (PRA) experiment before encounter⁴ at a radial distance of $160 R_J$ ($1 R_J$, Jupiter's radius = 7.14×10^4 km). The lower panel shows nKOM as narrow (black) bands at ~ 100 kHz; the separate emissions at higher frequencies are jovian hectometric radiation. The upper panel illustrates the sense of polarization of the emissions; for nKOM, black indicates left-handed and white, right-handed. When the north magnetic pole of Jupiter tilts towards V1 (spacecraft in the north magnetic hemisphere) nKOM is left-hand polarized. The polarization is the opposite when V1 is in the south magnetic hemisphere.

There have been earlier suggestions that nKOM derives from the Io plasma torus^{2,4,10}. The electron plasma frequency f_{pe} and upper hybrid frequency f_{UH} (see later) in the torus at $8\text{--}9 R_J$ are ~ 100 kHz, similar to the frequency of nKOM. Also, long-lived nKOM source regions sometimes showed a lag from co-rotation by 3–5%, taken to be suggestive of mass-loading effects at distances beyond the Io orbit⁴. Further, the observation of nKOM relatively close to the torus restricts the position of the source to the outer torus because of the shadowing by the torus of waves propagating from locations closer to Jupiter⁴. Jones¹⁰ proposed that jovian kilometric radiation was produced by mode conversion, via a radio window, of electrostatic upper hybrid (ESUH) waves to O mode electromagnetic radiation. It has since been shown⁸ that bKOM is well explained by such a conversion at sources very near the magnetic equatorial plane in the Io torus at $6\text{--}13 R_J$. The method used in that study of bKOM is applied here to nKOM.

The linear 'window' conversion theory requires the plasma density gradient ∇N_e in the source region to be nearly normal

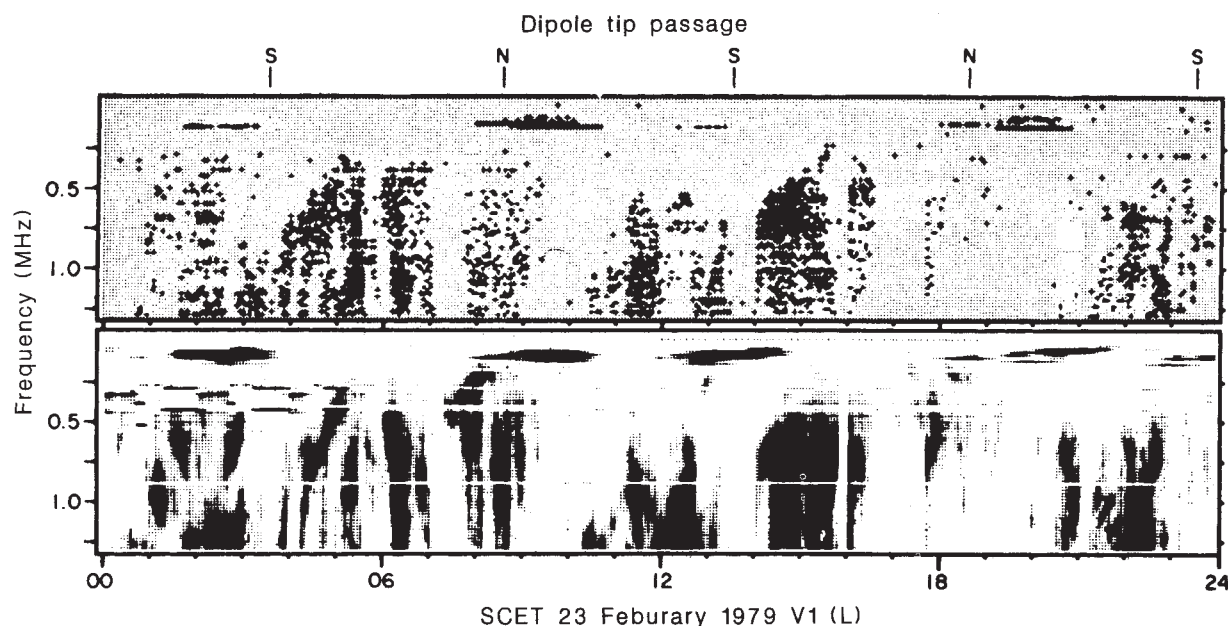


Fig. 1 Spectrograms of nKOM received on Voyager 1 when inbound at a distance of $160 R_J$ from Jupiter⁴. The nKOM emissions are seen in the lower panel as relatively narrow dark lines at ~ 100 kHz. The upper panel shows the corresponding nKOM polarization, black being left-handed and white, right-handed.

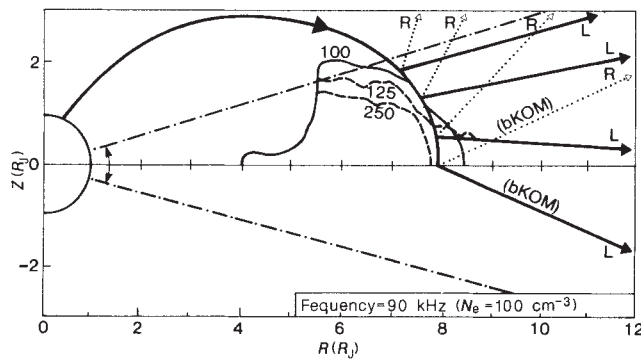


Fig. 2 A model¹² 100 cm⁻³ isodensity contour (full line) within the Io plasma torus, together with 125, 250 cm⁻³ contours (dashed lines) obtained from *in situ* measurements of the torus by V1¹¹. The range of magnetic latitudes covered by V1 and V2 inbound and outbound is bounded by the chained lines. The O mode rays from four hypothetical sources on the magnetic field line are shown assuming the field line to be an isodensity contour (with $N_e = 100$ cm⁻³) and assuming emission has occurred by linear 'window' mode conversion. The rays are labelled R (right) and L (left) to denote the polarization that would be observed by the Voyager spacecraft. The radiation from the equatorial source is bKOM^{8,9}. R is the radial distance in jovian radii (R_J) and Z is distance normal to the equatorial plane.

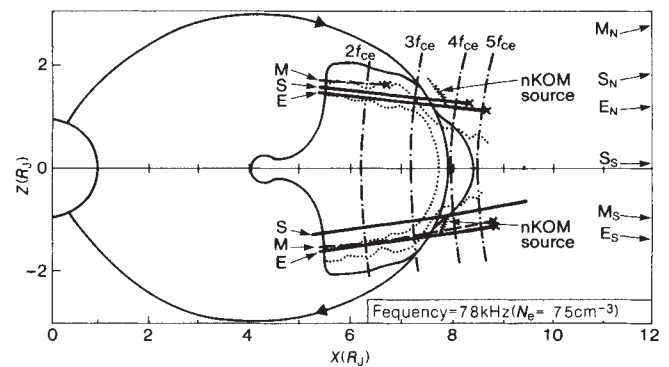


Fig. 3 Remote-sensing results of 78-kHz nKOM sources overlaid on isodensity contours in the Io plasma torus. The heavy, full lines labelled S (start) and E (end) constitute possible source locations of the start and end of the emissions observed by V1. When the start and end of the northern emission were observed, V1 was at the latitudes denoted by S_N , E_N respectively. The corresponding latitudes when observing the southern source limits are denoted S_S and E_S . M_N and M_S show latitudes in between when the possible source locations marked M were computed. The nearly vertical lines depict where the cyclotron frequency $f_{ce} = 78$ kHz/ n where $n = 2, \dots, 5$. The probable nKOM sources lie at the locations shown.

to the magnetic field vector $\mathbf{B}_0$. A result of conversion is that the ordinary (O) mode electromagnetic radiation is beamed through the magnetospheric cavity in the plane containing ∇N_e and $\mathbf{B}_0$, at angle $\beta = \pm \arctan (f_{pe}/f_{ce})^{1/2}$ with respect to $\mathbf{B}_0$, where plasma frequency $f_{pe} = 9 N_e^{1/2}$ Hz, cyclotron frequency $f_{ce} = 28 B_0$ Hz, plasma density is N_e m⁻³ and magnetic field strength is B_0 nT. If the sources are confined to the magnetic equatorial plane the radiation is beamed with respect to that plane, at angles $\alpha = \pm \arctan (f_{ce}/f_{pe})^{1/2}$, one beam to the south, the other to the north.

Figure 2 depicts Jupiter together with three isodensity contours in the Io plasma torus. The contours labelled 125 and 250 cm⁻³ are taken from *in situ* measurements¹² on V1 and the 100 cm⁻³ contour is taken from a computer model¹³ based on the measurements. A model magnetic field line including current sheet component¹⁴ is also shown as are the extremities of magnetic latitude attained by the Voyager probes inbound and outbound.

Assume the plasma density along the magnetic field line is 100 cm⁻³, that is, $f_{pe} = 90$ kHz; f_{ce} at the equator at $\sim 8 R_J$ is ~ 19 kHz and if 'window' conversion were to occur, the emitted radiation would be beamed at angles $\pm 25^\circ$ with respect to the equatorial plane (see Fig. 2). In a certain range of magnetic latitudes in the southern hemisphere the spacecraft would observe the southern left-hand polarized beam and at similar latitudes in the northern hemisphere they would observe the northern, right-hand polarized beam. This was the model proposed to explain the observed beaming of bKOM⁸.

Consider the possibility that the same type of conversion could occur at other points along the magnetic field line. In Fig. 2, hypothetical 90-kHz sources are placed at 5° , 10° , 15° latitude and, taking into account the variation in f_{ce} , the beam directions computed are shown. Similar beams would exist for southern sources. For distant spacecraft in the northern and southern hemispheres to observe respectively left-hand and right-hand polarized radiation, the sources must lie above $\sim \pm 8^\circ$ latitude. If the plasma density along the field line had a value other than 100 cm⁻³ this minimum source latitude would change because the beaming angle changes with f_{pe} .

The relation $\beta = \pm \arctan (f_{pe}/f_{ce})^{1/2}$ allows the use of a form of remote sensing to determine the possible positions of nKOM sources. The technique has been described elsewhere^{8,9,13} and

was used on bKOM to obtain torus plasma density profiles at the equatorial plane which are in remarkable agreement with the *in situ* profile obtained by V1^{8,9}. A brief summary of the remote-sensing technique is given below.

Given that the radiation is emitted at $f = f_{pe}$ and is beamed relative to the source magnetic field at an angle which depends on f_{pe} and f_{ce} , one may determine, for a given magnetic field model, where an observer has to be located to receive the radiation. Conversely, knowing the magnetic coordinates of the observer one can determine possible positions of the sources in the model magnetic field. The technique has been applied to nKOM observed by V1 using a composite magnetic field model including the current sheet contribution¹⁴.

The magnetic coordinates of V1 were determined when the various frequencies constituting the nKOM in Fig. 1 started and ended. If the nKOM is produced by linear conversion, these times would correspond to V1's entry into, and exit from, the beam. Source latitudes up to the limit of the Io torus were assumed and the radial positions of the sources satisfying the remote-sensing equation were computed. The results for $f = 78$ kHz are shown in Fig. 3.

When V1 was moving south at latitude 0.9° at 01:30 Spacecraft Event Time (SCET) on 23 February 1979 (see Fig. 1) it entered the 78 kHz beam from the southern source; the spacecraft latitude is indicated by the arrow marked S_S (Fig. 3, right). The end of the event occurred when V1 was at its most southerly latitude, -6.4° (E_S in Fig. 3). The emission observed between these two locations consisted of a fairly broad peak, intensity increasing and decreasing fairly smoothly as the spacecraft moved from S_S to E_S ; the maximum signal was recorded at the position marked M_S although it was little greater than other maxima within the main emission. The nearly horizontal lines in the southern half of the Io torus labelled S, M and E are the corresponding loci of possible source positions as obtained by remote sensing.

The same procedure was applied to many other events and the source loci for the northern beam for the event before that above are shown in the northern half of the torus. The maximum signal was recorded near the spacecraft's most northerly position (M_N) which may indicate that V1 had not reached the maximum within the beam. For both northern and southern beams, the possible source locations in the vicinity of the Io torus are much

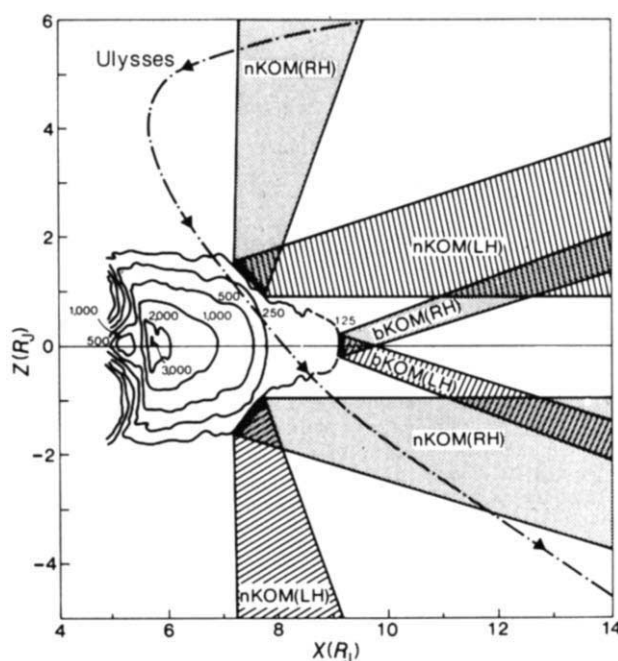


Fig. 4 Summary picture showing examples of the proposed nKOM and bKOM⁸ 100 kHz sources. The various beams are sketched and their polarizations indicated. The chained line shows the predicted orbit of the spacecraft Ulysses, which should allow the first observation of the northern hemisphere right-hand polarized beam of nKOM. (Note that because the time of arrival and thus the CML of Ulysses is as yet unknown, this trajectory has been drawn in jovigraphic coordinates whereas the remainder of the diagram is in jovimagnetic coordinates.)

as expected from the approximate beam positions computed for 90 kHz (Fig. 2). Taking into account that the crosses on the loci in Fig. 3 indicate the maximum radial distance at which, according to the theory, a source is visible to V1, and that the torus as determined *in situ* by V1 is as shown, the most likely sources of nKOM lie at $\sim 8 R_j$ at $8-12^\circ$ and $7-10^\circ$ in the northern and southern hemispheres respectively. In these regions the source-ESUH waves at 78 kHz (at $f_{UH} = (f_{pe}^2 + f_{ce}^2)^{1/2} = 78$ kHz) lie between the third and fourth (or fourth and fifth) harmonics of f_{ce} (see Fig. 3). It is well known that the intensity of ESUH waves maximizes at $\sim (n + 1/2)f_{ce}$ and that they suffer cyclotron damping at nf_{ce} . Remote sensing using other frequencies (97–154.8 kHz) for these and other nKOM events show similar results.

Near the suggested source locations the magnetic field line appears nearly parallel to the isodensity contours as modelled from *in situ* measurements; as already stressed, this is a condition required by the linear mode conversion mechanism (∇N_e normal to $\mathbf{B}_0$). With the available models, at no other point in the outer torus does this condition appear to be satisfied except near the equatorial plane where the bKOM sources are believed to be located^{8,9}. If nKOM sources are in the positions computed here, many of the radiation's characteristics can be readily explained.

The lower intensity of nKOM compared to bKOM may be attributed to the lower intensity of ESUH waves away from the equatorial plane. Although ESUH waves were detected over a relatively large latitude range in the Io torus, they were not as intense as those seen near the magnetic equator¹⁵.

Given the existence of the ESUH waves, the frequency of nKOM would be expected to be the plasma frequency where there is a density gradient and where the condition of normality is satisfied. It has been shown that the plasma frequency in the suggested source locations is ~ 100 kHz which is the approxi-

mate frequency of nKOM. Variations of this frequency would occur due to temporal and spatial variations of plasma density in the corresponding torus region. The total bandwidth of nKOM would be expected to correspond to the range of upper hybrid and plasma frequencies in the source region. If this bandwidth were larger than f_{ce} , nKOM would itself consist of one or more narrow band emissions corresponding to $f_{UH} \approx (n + 1/2)f_{ce}$. It appears that at the source, $f_{ce} \approx 20$ kHz which would result in a minimum bandwidth of < 20 kHz in agreement with observation. It should be noted that the PRA receiver² consists of a number of 1-kHz bandwidth filters spaced 19.2 kHz apart, set at frequencies 20.4, 39.6 kHz and so on. Therefore, any fine structure in nKOM with a 20-kHz spacing would not have been resolved. Indeed it is conceivable that on occasions no signal would have been recorded in any of the PRA receiver channels since the emission's frequency bands would have lain between filters.

The computed source locations also explain the smooth temporal nature of nKOM in contrast with the sporadic character of bKOM. Even allowing for variations in f_p in the nKOM source regions, the source-ESUH waves have a fairly wide frequency range (≤ 20 kHz) in which to exist without being damped. It is plausible that the 78-kHz emission region in Fig. 3, for example, remained within the $3f_{ce}-4f_{ce}$ band during the period of observation. The sources of bKOM at corresponding frequencies lie in regions^{8,9} where $f_{UH}/f_{ce} \sim 10$, so that relatively small fluctuations in plasma density (three times smaller in percentage terms) would cause the frequency f_{UH} often to wander between cyclotron harmonic bands resulting in the disappearance of bKOM when $f_{UH} = nf_{ce}$ and its reappearance when $f_{UH} \neq f_{ce}$.

The polarization of nKOM observed by the Voyager probes has already been accounted for. There are cases when the polarization reversal does not occur exactly at the magnetic equator^{4,5}; one such case is shown in Fig. 3 where the southern, right-handed source is first observed just north of the equatorial plane. This could be due to asymmetries in the plasma torus, either locally or due to the small angle ($\sim 3^\circ$) between the magnetic equatorial plane and the centrifugal equatorial plane, the latter being considered to be the symmetry plane of the Io plasma torus¹⁶.

One observation not explained by the present work is the apparent lag of some long-lived nKOM sources relative to co-rotation and the absence of such an effect in bKOM sources which, because of their greater radial distance, should be more susceptible to mass-loading. It may be that the effect is easier to observe in nKOM because of its smoother nature. It may also be an artefact due in some way to the PRA experiment's limitation in being able to observe only certain nearly harmonically related frequencies. The latter may in fact explain why some nKOM sources disappear but later reappear at the correct phase, their disappearance possibly being due to the nKOM frequencies falling between PRA filters as explained earlier. This implies that either the source gradient is changing or that it is moving radially, either out or in. The latter would change the nKOM beaming, an inward motion causing the beams to be dipped to lower latitudes. This would appear as a shift in CML for Voyager observations, but this shift could be either forwards or backwards. Studies are underway using nKOM events showing a co-rotation lag in an endeavour to obtain an explanation of this phenomenon.

Figure 4 is a summary of the proposed locations of nKOM and bKOM (78 kHz) sources, and the polarized beams of radiation expected from them. Also shown is the proposed trajectory of Ulysses, a spacecraft which in the next decade is intended to investigate the Sun and solar wind at high ecliptic latitudes. Its Radio and Plasma Wave Experiment¹⁷ should allow the first observations of the high-latitude northern hemisphere, right-hand polarized nKOM beam.

The results presented here show that nKOM may be produced by linear 'window' conversion. This would add it to the family

of emissions, bKOM from Jupiter, myriametric radiation from the Earth^{11,18} and Saturn¹⁹, which are all apparently produced by the same generation mechanism, a process which may well be universal with applications to solar and other astrophysical radio sources.

I thank Y. Leblanc for providing high resolution PRA data, J. Connerney for the magnetic field model, and K. G. Budden and F. Bagenal for useful discussions.

Received 5 February; accepted 10 April 1987.

1. Warwick, J. W. *et al. Science* **204**, 991-995 (1979).
2. Warwick, J. W. *et al. Science* **204**, 995-998 (1979).
3. Scarf, F. L. *et al. Science* **204**, 991-995 (1979).
4. Kaiser, M. L. & Desch, M. D. *Geophys. Res. Lett.* **7**, 389-392 (1982).
5. Daigne, G. & Leblanc, Y. *J. geophys. Res.* **91**, 7961-7969 (1986).
6. Desch, M. D. & Kaiser, M. L. *J. geophys. Res.* **85**, 4248-4256 (1980).
7. Leblanc, Y. & Daigne, G. *J. geophys. Res.* **90**, 12073-12080 (1985).
8. Jones, D. *Nature* **324**, 40-42 (1986).
9. Jones, D. & Leblanc, Y. *Ann. Geophys.* **5A**, 29-38 (1987).
10. Jones, D. *Nature* **228**, 225-229 (1980).
11. Bagenal, F. *et al. J. geophys. Res.* **90**, 1755-1757 (1985).
12. Divine, N. & Garrett, H. B. *J. geophys. Res.* **88**, 6889-6903 (1983).
13. Jones, D. *J. geophys. Res.* **88**, 158-166 (1983).
14. Connerney, J. E. P. *et al. J. geophys. Res.* **86**, 8370-8384 (1981).
15. Barbosa, D. D. & Kurth, W. S. *J. geophys. Res.* **85**, 6729-6742 (1980).
16. Dessler, A. J. *Physics of the Jovian Magnetosphere* (ed. Dessler, A. J.) 498-504 (Cambridge University Press).
17. Stone, R. G. *et al. The International Solar Polar Mission—Its Scientific Investigations* (eds Wenzel, K. P., Marsden, R. G. & Battrick, B.) 185-207 (ESA Publications Division, ESTEC, The Netherlands, 1983).
18. Jones, D. *et al. Nature* (submitted).
19. Jones, D. *Nature* **306**, 453-456 (1983).

Latitudinal variation in oxygen-18 of atmospheric CO₂

Roger J. Francey* & Pieter P. Tans†

* CSIRO Division of Atmospheric Research, Aspendale, Victoria 3195, Australia

† Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado/NOAA, Boulder, Colorado 80309, USA

This report provides information on a potentially important new atmospheric tracer of large-scale behaviour at the Earth's surface, the oxygen isotope composition of CO₂. We use measurements on flask air collected from Alaska, Hawaii, Samoa, Tasmania, coastal Antarctica and the South Pole. Recently, we examined 1982-84 measurements of $\delta^{18}\text{O}$ in CO₂ extracted *in situ* from marine air at Cape Grim, Tasmania¹. Here we report on a comparison of Cape Grim flask and *in situ* data that gives a measure of precision and serves to demonstrate a marked improvement over previous infrequent measurements. While previous data^{2,3} suggests a north-south gradient, our flask data establish a strong, asymmetric meridional gradient. We argue that this reflects the oxygen isotope ratio in ground water, via mechanisms involving gross catalysed CO₂ exchange with leaf (and possibly soil) water. Very large CO₂ fluxes are involved, of the order of 200 Gt carbon (C) yr⁻¹.

Sampling sites have been chosen for their access to large well-mixed air masses, remote from human influence. Monthly flask sample pairs have been collected in clean air conditions. Filling a flask involves venting off clean, dry, storage air, flushing for 15-20 min at 5-10 l min⁻¹ with a metal bellows or Teflon diaphragm pump which draws air through a magnesium perchlorate drying tower, then pressurizing to 100 kPa above ambient. The process is immediately repeated for the second flask. On return to CSIRO, the flasks were first analysed for mixing ratios of CO₂, CO, CH₄, N₂O, CCl₃F and CCl₂F₂ by gas chromatography. Then CO₂ was extracted cryogenically for mass spectrometer analyses of $\delta^{13}\text{C}$ and $\delta^{18}\text{O}$.

Figure 1 compares $\delta^{18}\text{O}$ measurements on Cape Grim air

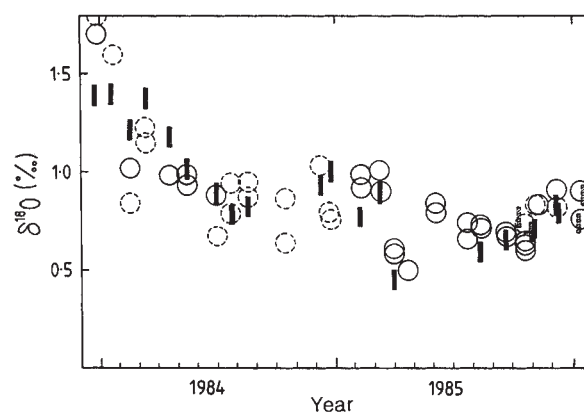


Fig. 1 A comparison of $\delta^{18}\text{O}$ in marine atmospheric CO₂ using two different methods. The bars represent *in situ* CO₂ extraction from Cape Grim 'baseline' air. The circles are from whole-air collections in pressurized 5 l glass flasks. Dashed circles indicate minor reported technique anomalies or minor flask contamination of other trace species. When contemporary *in situ* analyses are not available, dashed bars show analyses on nearby dates. ($\delta^{18}\text{O}$ are with respect to PDB-CO₂.)

from the *in situ* CO₂ extraction program⁴ with those from samples collected in flasks at the same time.

There are differences between the *in situ* technique at Cape Grim and the flask technique which are significant in terms of detecting possible systematic error in $\delta^{18}\text{O}$. The Cape Grim *in situ* method uses cryogenic pre-drying of the air stream compared to the chemical drying in the flasks; the *in situ* values represent a 2 h integration of air encompassing 5-10 min integrations for the flasks; dynamic flow conditions during CO₂ extraction are quite different, with the flask pressures dropping from 100 kPa above ambient to vacuum; flask samples are generally analysed in the mass spectrometer weeks to months after contemporary *in situ* samples.

In view of the significant differences in technique, the excellent agreement is interpreted as substantial support for the integrity of both methods. In particular, the emphasis on drying flask air before storage, which distinguishes this programme from previous programmes measuring the $\delta^{18}\text{O}$ of atmospheric CO₂, is considered significant. It is thought that this minimizes the opportunity for CO₂ hydration and subsequent oxygen exchange. In Fig. 2 selected data from all stations are plotted.

A prominent feature of the data is the existence of a meridional gradient: the $\delta^{18}\text{O}$ values are relatively constant in southern latitudes (a fact not suggested in previous data^{2,3}) but become steadily lower going north from Cape Grim. The gradient is strongest between Mauna Loa and Pt Barrow. A seasonal cycle may exist but is not very prominent in the data. There is also a strong suggestion of interannual variability that is coherent among the stations, with values in 1984 higher than those in 1985. Both major features are summarized in Table 1.

The isotope exchange flux that is needed to maintain an average gradient of 1.2‰ between Barrow and Mauna Loa against vigorous atmospheric mixing is enormous. An order of

Table 1 Weighted annual mean $\delta^{18}\text{O}$ values (‰) for atmospheric CO₂ as a function of latitude for 1984 and 1985

Station	Latitude	1984	1985
SPO	90° S	0.80	0.75
MAW	67° S		0.82
CGO	41° S	0.96	0.75
SMO	14° S	0.64	0.35
MLO	20° N	0.27	0.01
BRW	71° N	-0.96	-1.10

magnitude calculation illustrates this. Consider the Northern Hemisphere as divided into a northern and southern half, each containing 175 Gt C (1 Gt C = 10^{15} g of carbon) in the form of CO_2 . Assume that the exchange time between the two parts is as long as one-quarter of a year. Assume also that the carbon dioxide in both halves is trying to equilibrate its oxygen with a source that is 10% lower in the north than in the south⁵. Then the fraction of the CO_2 in the Northern Hemisphere that exchanges oxygen with the source is $4 \times 1.2 / (10 - 1.2) = 0.55$, amounting to almost 200 Gt C yr^{-1} . For comparison, the total net primary production of all plants on earth is estimated to be 40–70 Gt C yr^{-1} (ref. 6) and the total global exchange of CO_2 across the air-sea interface is around 90 Gt C yr^{-1} (ref. 7).

We do not think that isotope exchange with ocean water or cloud droplets⁸ can explain the observed gradient. Mills and Urey⁹ studied the exchange of oxygen isotopes between CO_2 and liquid water. They found that hydration of the CO_2 molecule is necessary for exchange to occur and also that hydration proceeds slowly, with a time constant of 30 s at 25 °C. Therefore, isotope exchange of CO_2 with ocean water is limited by the rate at which CO_2 can diffuse across the liquid boundary film, and the air-sea CO_2 flux of around 90 Gt C yr^{-1} (ref. 7) is too small to maintain a gradient as large as that observed in the Northern Hemisphere. A second compelling argument is that the isotope signature of CO_2 equilibrated with cold ocean water at the poles and warm water at the Equator is different from what we observe. The maximum $\delta^{18}\text{O}$ difference in equilibrated CO_2 (without atmospheric mixing) would only be a few ‰ with a minimum in $\delta^{18}\text{O}$ at the Equator.

The surface area of clouds is enormous so that the entry of CO_2 molecules into cloud water is not rate-limiting⁸. But an individual CO_2 molecule must accumulate ~ 30 s of residence time in liquid to permit hydration and isotope exchange. In a typical cloud droplet of 10 μm diameter, the average distance (x) to the nearest surface is one-third of the droplet radius, or 1.7 μm . The time to travel this distance by diffusion is

$$t = x^2 / 4D$$

which, for a diffusion coefficient of CO_2 in water of $D = 2 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$, is 0.4 ms. The fraction of the total time that CO_2 molecules spend in liquid water droplets thus depends on the amount of liquid water (w) in the atmospheric column, and the solubility (s) of CO_2 . For $s = 0.05 \text{ mol l}^{-1} \text{ atm}^{-1}$ (at 12 °C) the dissolved CO_2 content of the cloud water is $18 \mu\text{mol l}^{-1}$. For example, the annual average liquid water content of clouds over Moscow is $\bar{w} = 1.4 \text{ g m}^{-3}$, with an average cloud depth of 350 m (ref. 10), giving a column density of $5 \times 10^{-5} \text{ kg cm}^{-2}$. Using this as a first estimate, the fraction of CO_2 dissolved in cloud water relative to all atmospheric CO_2 is then

$$\frac{18 \times 10^{-6} \times 5 \times 10^{-5}}{340 \times 10^{-6} \times 10^3 / 29} = 7 \times 10^{-8}$$

Thus, to accumulate 30 s of residence in liquid water an average CO_2 molecule requires nearly 14 years. Most cloud droplets are much colder than 25 °C and hydration takes even longer at lower temperatures. In support of these estimates, the Cape Grim *in situ* measurements (several per month over 4 years) were found to be quite insensitive to meteorological conditions (for example, rain and fog). In addition, the asymmetry and sense of the observed latitudinal gradient is difficult to reconcile with exchange with atmospheric water.

The observed latitudinal distribution is reminiscent of that due to anthropogenic emissions. The combustion of fossil fuels combines carbon with atmospheric oxygen that is isotopically lighter than CO_2 by about 20‰¹¹. If we assume no isotope fractionation between CO_2 and H_2O produced in the combustion, then the fossil fuel should have a -20% isotope signature. If we further assume an atmospheric exchange time between the hemispheres of one year, then the fossil CO_2 should result

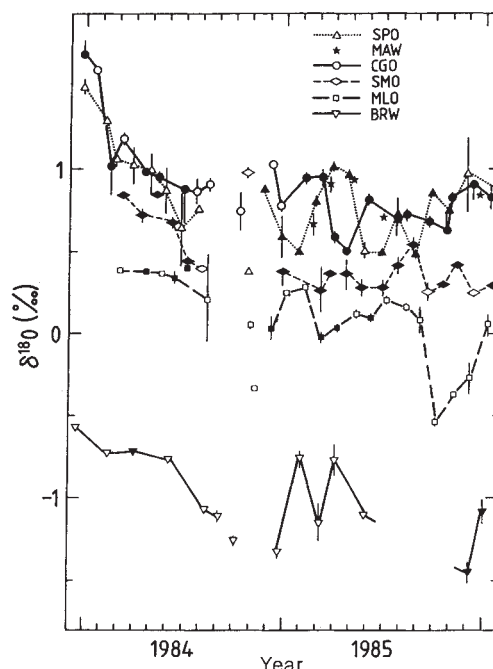


Fig. 2 Mean $\delta^{18}\text{O}$ obtained from CO_2 in the 5 l pressurized glass flasks collected from the global network (see text). Open symbols are used to indicate suspect data on the basis of minor anomalies. Vertical bars span the range of values for multiple flask samples. The sampling sites (with latitudes) are South Pole, Mawson, Cape Grim, Samoa, Mauna Loa and Barrow.

in an interhemispheric difference of about $5/350 \times 20 = 0.3\%$ because there is about 350 Gt C per hemisphere. Such a gradient is not nearly enough to account for our observations.

We consider the exchange of CO_2 with atmospheric sources of oxygen other than water, as a very unlikely cause of the observed gradient. None of these molecules has fluxes into and out of the atmosphere that seems remotely capable of sustaining their own isotope gradient while driving the enormous exchange rate with CO_2 necessary to maintain the observed ^{18}O in CO_2 gradient. Atmospheric oxygen itself does not seem to exchange at all; it is not in isotope equilibrium with either water or CO_2 ; its isotopic composition is constant everywhere¹² and seems to be the result of plant respiration. A possible candidate for large exchange fluxes with CO_2 would be direct ^{18}O exchange with water in the vapour phase. Because of the large amounts of water vapour present in the atmosphere, in the neighbourhood of 3×10^{17} molecules cm^{-3} in the boundary layer, a time constant for isotope equilibration of CO_2 to water vapour of $3 \times 10^{-8} \text{ s}^{-1}$, or about 1 yr^{-1} , implies a rate constant which is very small, of the order of $10^{-25} \text{ cm}^3 \text{ s}^{-1}$. This mechanism appears to be excluded on the basis of preliminary isotope exchange experiments in the laboratory, as well as the asymmetrical nature of the observed ^{18}O in the CO_2 gradient.

Another source of oxygen isotopes that seems potentially capable of producing the observations of $\delta^{18}\text{O}$ in CO_2 is surface water on the great continental land masses of the Northern Hemisphere. The precipitation is progressively depleted of ^{18}O at higher latitudes and deeper into the continental interiors¹⁰. For instance, the $\delta^{18}\text{O}$ of the precipitation in Northern Canada is between -15 and -30% . The question is how this water communicates with the CO_2 .

An obvious mechanism is related to the photosynthesis of plants. When they open their stomata, CO_2 diffuses into the intracellular spaces and from there into the leaf water. The concentration of CO_2 in the intracellular spaces is about two-thirds of the atmosphere. This means that one-third of the CO_2 entering the leaves is fixed during photosynthesis as gross primary production (GPP), while two-thirds diffuses back out.

Although the length of stay inside the leaf of a CO_2 molecule is ~ 1 s, we expect its oxygen to be equilibrated fully with the water because of the ubiquitous presence of the enzyme carbonic anhydrase¹³. The catalysis of the hydration of CO_2 by carbonic anhydrase, with maximal turnover number near 10^6 s^{-1} , is one of the fastest known enzymic reactions¹⁴. GPP is estimated to be two to four times as large as net primary production (NPP)¹⁵, where NPP is equal to GPP minus autotrophic respiration. So we have a flux of CO_2 , that equilibrates with water, of about $50 \times 2 \times (2-4) = 200\text{--}400 \text{ Gt C yr}^{-1}$.

Thus exchange with leaf water provides a reasonable mechanism for $\delta^{18}\text{O}$ in CO_2 modification on the space- and timescales observed, however there are still problems to be resolved. First, one would expect the $\delta^{18}\text{O}$ gradient to be seasonal; while small amplitude seasonality is apparent at Cape Grim¹, a conclusion concerning the Northern Hemisphere stations must await further data. Second, the water in the leaves of photosynthesizing plants is significantly enriched in ^{18}O because H_2^{16}O evaporates more readily than H_2^{18}O .

Mass balance considerations in a simple steady-state model of evapotranspiration show that the isotope enrichment of the water in the plant cells is:

$$\delta_{\text{cw}} - \delta_{\text{g}} = \varepsilon_{\phi} - (1-h)(\varepsilon_{\text{d}} + \varepsilon_{\text{e}}) + h(\delta_{\text{v}} - \delta_{\text{g}})$$

where δ_{cw} , δ_{g} , δ_{v} are the isotope ratios in cell water, ground water and atmospheric water vapour respectively, h is the relative humidity of the air. The ε s are expressed in p.p.m. and are related to the isotope fractionation factors α , by $\alpha_{\phi} = 1 + \varepsilon_{\phi}$ and so on, α_{ϕ} , α_{e} and α_{d} refer to the equilibrium isotope fractionation, respectively, of cell water with respect to pure water (due to solutes and matrix effects), of vapour with respect to pure water, and the kinetic fractionation of transport through the stomates and beyond. A similar expression has been obtained independently by Farquhar (personal communication). If the kinetic effects are dominated by molecular diffusion, we have $\varepsilon_{\text{d}} = -16\%$ for deuterium and $\varepsilon_{\text{d}} = -31\%$ for oxygen-18, while $\varepsilon_{\text{e}} = -78\%$ and $\varepsilon_{\text{e}} = -10\%$ (at 20°C) respectively. If $\delta_{\text{v}} = \delta_{\text{g}}$ then we find 2.3 for the slope of the evapotranspiration line in a plot of deuterium against oxygen-18, which is exactly what Epstein *et al.* observe for terrestrial plants¹⁶. Using our formula for the enrichment, with a relative humidity of 0.7 during photosynthetic uptake, we find from global model calculations that $\delta^{18}\text{O}$ of CO_2 would be more enriched than what is observed by 1–2%. The discrepancy would be greater if the relative humidity is lower during the time that the stomata are open. On this basis we can speculate that there must be another exchange mechanism that balances the seasonality and the evaporative enrichment of the plants and possibly contributes to the $\delta^{18}\text{O}$ gradient in CO_2 . A likely candidate is the isotope exchange between water and CO_2 in soils that can be catalysed either by the presence of carbonic anhydrase in soils or by inorganic catalytic sites such as metal oxides¹⁷. Seasonality might be provided in this scheme by variable snow cover shutting off the ground from effective contact with the atmosphere.

The flask sampling program was funded by the Australian National Energy Research, Development and Demonstration Council. We thank Graeme Pearman, Paul Fraser, Geoff Richards, Anne Becker, Helen Goodman, Scott Coram and Ian Weeks for their help. We thank Graham Farquhar for valuable discussions and the management and staff of the Australian Cape Grim Baseline Air Pollution Station, the NOAA/GMCC stations at Barrow, Mauna Loa, Samoa and South Pole, and the Australian Antarctic Research Expedition at Mawson for their assistance.

Note added in proof: Dr U. Siegenthaler has made useful comments and pointed out a reference by Friedli, H., Siegenthaler, U., Rauber, D. & Oeschger, H. *Tellus* (in the press). Our arguments concerning exchange with cloud water are supported. These authors also conclude that exchange with the biosphere must be important.

Received 13 October 1986; accepted 2 April 1987.

1. Tans, P. P., Francey, R. J. & Pearman, G. I. in *Baseline 83–84* (eds Francey, R. J. & Forgan, B. W.) 10–14 (Bureau of Meteorology/CSIRO Division of Atmospheric Research, Australia, 1986).
2. Keeling, C. D., Carter, A. F. & Mook, W. G. *J. geophys. Res.* **89**, 4615–4628 (1984).
3. Mook, W. G., Koopmans, M., Carter, A. F. & Keeling, C. D. *J. geophys. Res.* **88**, 10915–10933 (1983).
4. Francey, R. J. & Goodman, H. S. in *Baseline 83–84* (eds Francey, R. J. & Forgan, B. W.) 27–36 (Bureau of Meteorology/CSIRO Division of Atmospheric Research, Australia, 1986).
5. Joussaume, S., Sadourny, R. & Jouzel, J. *Nature* **311**, 24–29 (1984).
6. Pearman, G. I. & Hyson, P. *J. atmos. Chem.* **4**, 81–124 (1986).
7. Pearman, G., Hyson, P. & Fraser, P. *J. J. geophys. Res.* **88**, 3581–3590 (1983).
8. Bottinga, Y. & Craig, H. *Earth planet. Sci. Lett.* **5**, 285–295 (1969).
9. Mills, G. A. & Urey, H. C. *J. Am. chem. Soc.* **62**, 1019–1026 (1940).
10. Feigelson, E. M. *Radiant Heat Transfer in a Cloudy Atmosphere* (translated by Lederman, D.) (Israel Program for Scientific Translations, Jerusalem, 1973).
11. Kroopnick, P. & Craig, H. *Science* **175**, 54–55 (1972).
12. Dole, M., Lane, G. A., Rudd, D. P. & Zaukelies, D. A. *Geochim. cosmochim. Acta* **6**, 65–78 (1954).
13. Reed, M. & Graham, D. in *Progress in Phytochemistry* Vol. 7 (eds Reinhold, L., Marbone, J. & Swain, T.) 47–94 (Interscience Publishers, London, 1981).
14. Silverman, D. N. *Methods in Enzymology* **87**, 732–752 (1982).
15. Larcher, W. *Physiological Plant Ecology* (Springer, Berlin, 1975).
16. Epstein, S., Thompson, P. & Yapp, C. *J. Science* **198**, 1209–1215 (1977).
17. Newsome, D. S. *Catal. Rev. Sci. Engng* **21**, 275–318 (1980).

ENSO signal in continental temperature and precipitation records

R. S. Bradley*, H. F. Diaz†, G. N. Kiladis‡
& J. K. Eischeid‡

* University of Massachusetts, Amherst, Massachusetts 01003, USA

† National Oceanic and Atmospheric Administration, Environmental Research Laboratory, Boulder, Colorado 80303, USA

‡ Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado 80309, USA

Warm events in the tropical Pacific Ocean, associated with one extreme of the Southern Oscillation (generally known as El Niño–Southern Oscillation or ENSO events) are now known to be of global significance in perturbing the general circulation of the atmosphere^{1–8}. It is also apparent that episodes of enhanced cold-water upwelling ('cold events') have an impact on large-scale atmospheric circulation patterns^{9–11}. Here we examine the effect of warm and cold events on short-term fluctuations of continental surface air temperature and precipitation throughout the Northern Hemisphere using a newly compiled set of high-quality temperature and precipitation data^{12–13}. The data were in the form of gridded anomalies from a 1951–70 reference mean for temperature and a 1921–60 reference mean for precipitation^{13–14}.

A commonly used index of the Southern Oscillation is the normalized sea-level pressure difference between Tahiti and Darwin^{15,17}. With this index (SOI), large negative values correspond to warm events and large positive values to cold events. A five-month running mean of the index is shown in Fig. 1, together with a corresponding five-month running mean of average monthly temperature anomalies for the land areas of the latitude zone 0–25° N. This area includes sub-Saharan Africa, southern India, south-east Asia, central America and northern South America. It is clear that the temperature series exhibits periods of persistent positive anomalies separated abruptly from periods of persistent negative anomalies. Furthermore, the positive temperature anomalies correspond to times when the SOI is negative and vice versa. This is consistent with other studies (based on more limited data) which have reported evidence of a strong correlation between equatorial Pacific sea surface temperature (SST) and tropical tropospheric mean temperature (TTT), with SST leading TTT by about 6 months^{3,18–22}. Note also that the relative absence of periods of positive SOI in the past 10 years corresponds to a period of persistently high tem-

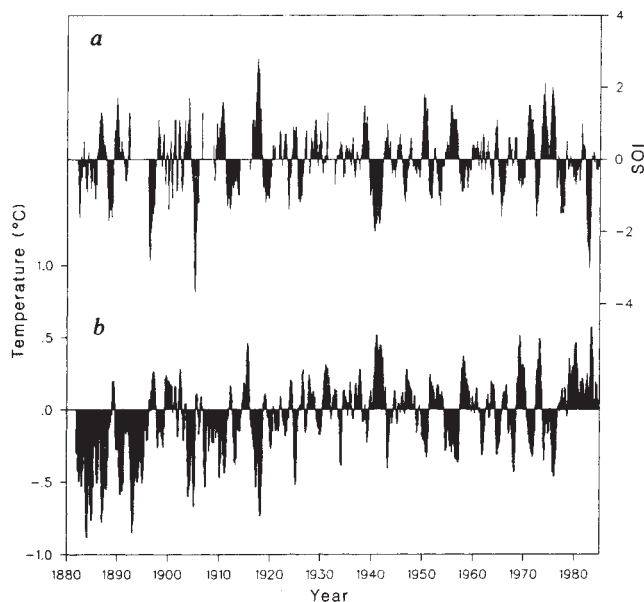


Fig. 1 *a*, Southern Oscillation index based on normalized series of Tahiti-Darwin sea-level pressure (both station sets normalized). Negative values correspond to warm (ENSO) events, the extremes of which are El Niño events; *b*, average temperature anomaly (with reference to 1951-1970) over continental areas 0-25° N based on a gridded data set¹³. Values are five-month running means.

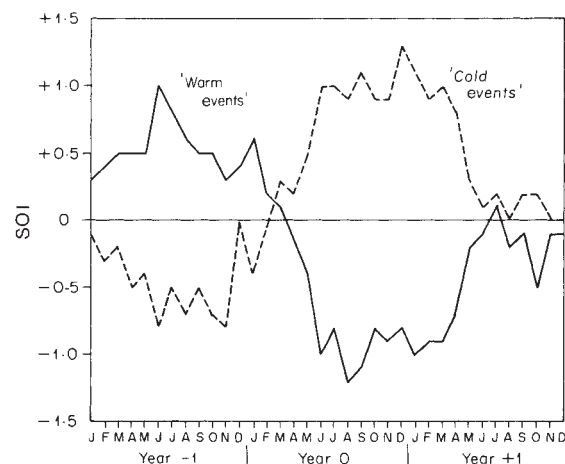


Fig. 2 Composite SOI sequence derived by averaging monthly values centred on the cold event and warm event years. The warm event years are: 1884, 1888, 1891, 1896, 1899, 1902, 1904, 1911, 1913, 1918, 1923, 1925, 1930, 1932, 1939, 1951, 1953, 1957, 1963, 1965, 1969, 1972, 1976. The cold event years are: 1886, 1889, 1892, 1898, 1903, 1906, 1908, 1916, 1920, 1924, 1928, 1931, 1938, 1942, 1949, 1954, 1964, 1970, 1973, 1975.

peratures. A regression of monthly SOI and monthly temperature anomalies over the period 1910-1980 shows a strong negative relationship ($r = -0.41$, $p = 0.01$), with maximum correlation when the SOI leads by five months.

To examine the time-dependent relationship between warm and cold events and continental surface temperatures at different latitudes, a superposed epoch analysis was performed²³. The selection of key dates was based on the Tahiti-Darwin pressure series and a separate index composed of sea surface temperature anomalies averaged from 4° N-4° S, 160° W to the South American coast²⁴. Based on these indices and previous studies^{10,24-26}, the starting year (year 0) of 23 warm events and 20 cold events was identified in the period 1881-1980 (see Fig. 2 legend). The SOI was then averaged by month, from January of the year preceding each event (year -1) to the December of the year following it (year +1). This produced a composite SOI time series for both ENSO extremes (Fig. 2). The SOI changes sign, on average, in April of the event year. Thus, April of each year shown in the legend of Fig. 2 was selected as the key month for superposed epoch analysis of temperature change over continental regions. Temperature anomalies were standardized (to accommodate differences in variance from summer to winter months) by dividing each monthly anomaly by the standard deviation of the respective month in the period 1881-1980. Because we are interested in the effects of ENSO events on prevailing temperatures at various times during each event, the superposed epoch analysis was carried out with respect to the mean of the 36 months preceding each warm and cold event. In this way the effects of warm and cold events during generally colder than average periods are comparable with their effects during generally warmer than average periods. Zonally averaged anomalies for each 5° latitude band from 5° N to 70° N were computed for the 36 months before and 36 months following each warm event. As a check, the same procedure was performed on two separate sets of both warm and cold event years; namely those before 1925 and after 1925. Similar results were obtained for each of these independent runs, confirming the reliability of the signals.

To assess the statistical significance of the anomalies a Monte Carlo approach was used. Four-hundred simulations of 20 randomly selected dates within the interval 1881-1980 were used to determine appropriate confidence levels. In general, the 5% significance level falls close to the $\pm 0.5\sigma$ composite mean level.

Figure 3 shows the composite warm and cold event time series of zonally-averaged temperature for the four latitude bands with the strongest signals (5° to 20° N). For warm events, approximately 12 months before April of year 0 an abrupt decline in temperature occurs in the tropics, reaching levels significantly below average from around July (year -1) to March (year 0) (Fig. 3*a-d*, solid curve). Temperatures then rise rapidly, reaching levels significantly above average from August through July of the following year, after which temperatures return to near normal. Maximum negative anomalies occur, on average, in the autumn and winter of year -1 and maximum positive anomalies in the winter and spring of year +1 (two to three months after the SOI composite minimum). However the switch from well below to well above average temperatures occurs very rapidly, in the course of only 3-4 months. The results of the same procedure applied to cold events are shown as the dashed curves in Fig. 3*a-d*. In the tropics the temperature anomalies clearly follow the opposite course, being well above average before April of year 0 and falling abruptly thereafter. Minimum temperatures occur in spring and early summer of year +1, 12 months or so after the SOI composite maximum. The composite responses to warm and cold events have a correlation coefficient of -0.80 ($p = 0.01$). This pattern of opposite trends is, in part, related to the fact that cold and warm events often follow each other closely in time (see Fig. 2 legend) and hence the superposed epoch analyses overlap in many instances. Nevertheless, the data clearly show that the change from one phase of the Southern Oscillation to another is reflected in relatively large shifts in the temperature anomaly pattern throughout the Northern Hemisphere tropics.

It has been shown that climate anomalies associated with the evolution of ENSO vary spatially as well as temporally²⁷⁻³⁰. Therefore, we repeated the analysis, this time focusing on selected continental regions from 0-30° N (Fig. 4). The three tropical regions selected cover the Americas (150° W-30° W), North Africa (30° W-60° E) and southern Asia (60° E-150° E). As in Fig. 3, all regions show above normal temperatures at the end of year 0 of warm events and below normal temperatures during cold events, with the opposite signal during year -1. This signal

Fig. 3 Superposed epoch analysis of temperature anomalies over continental areas for warm (solid) and cold (dashed) ENSO events. Temperature anomalies (from 1951 to 1970 mean) normalized by standard deviation of respective month for period 1881–1980. Anomalies expressed as departures from 36-month average before April of Year 0. Key years are shown in Fig. 2 legend. Dashed horizontal lines show levels of significance at the 55% level.

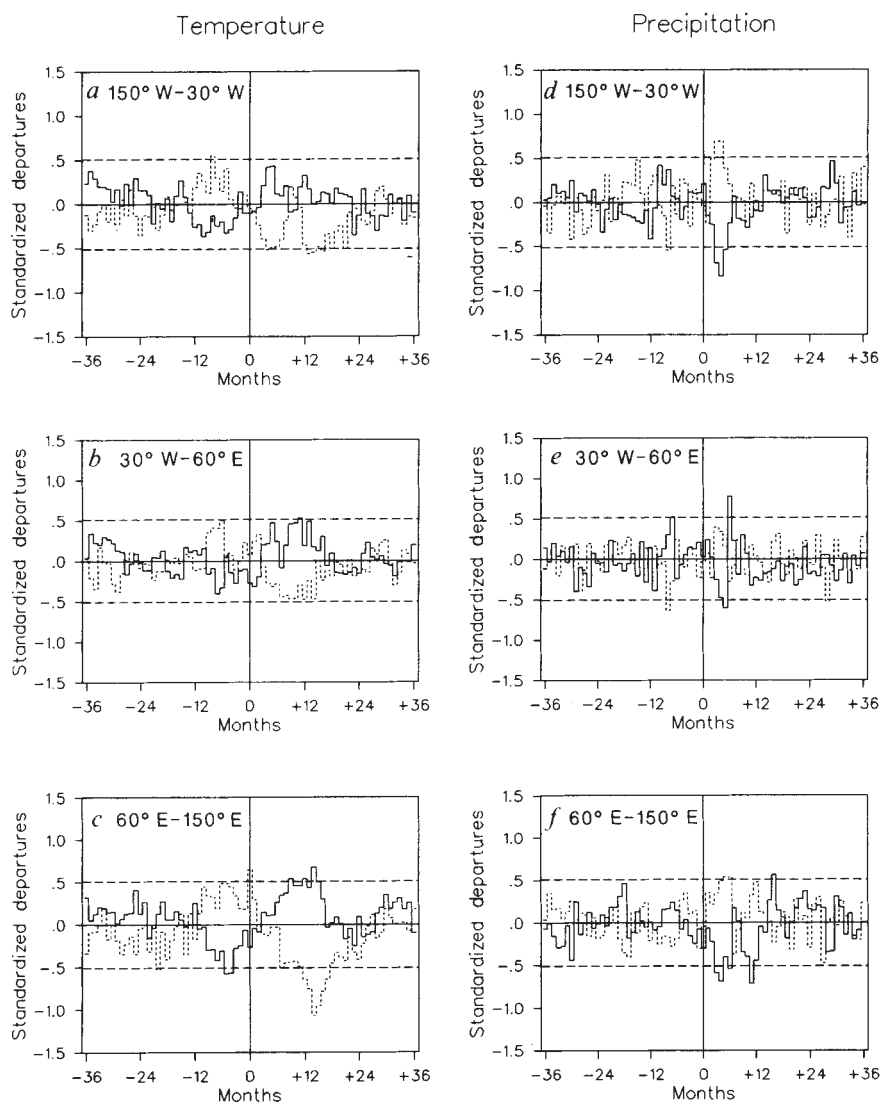
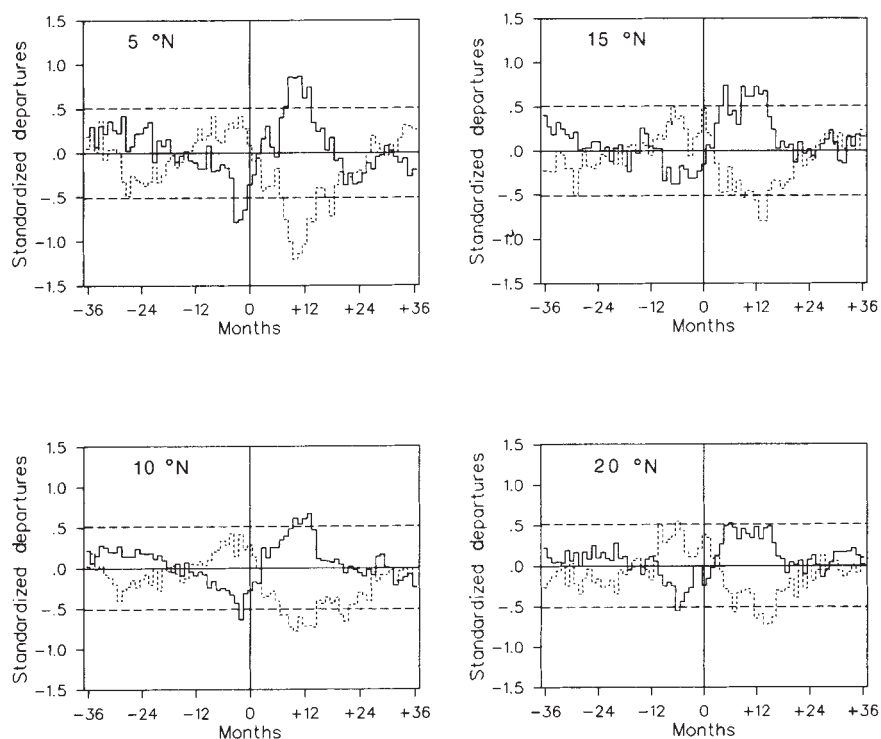


Fig. 4 As in Fig. 3, except for regional averages (0° – 30° N over the indicated longitudes).

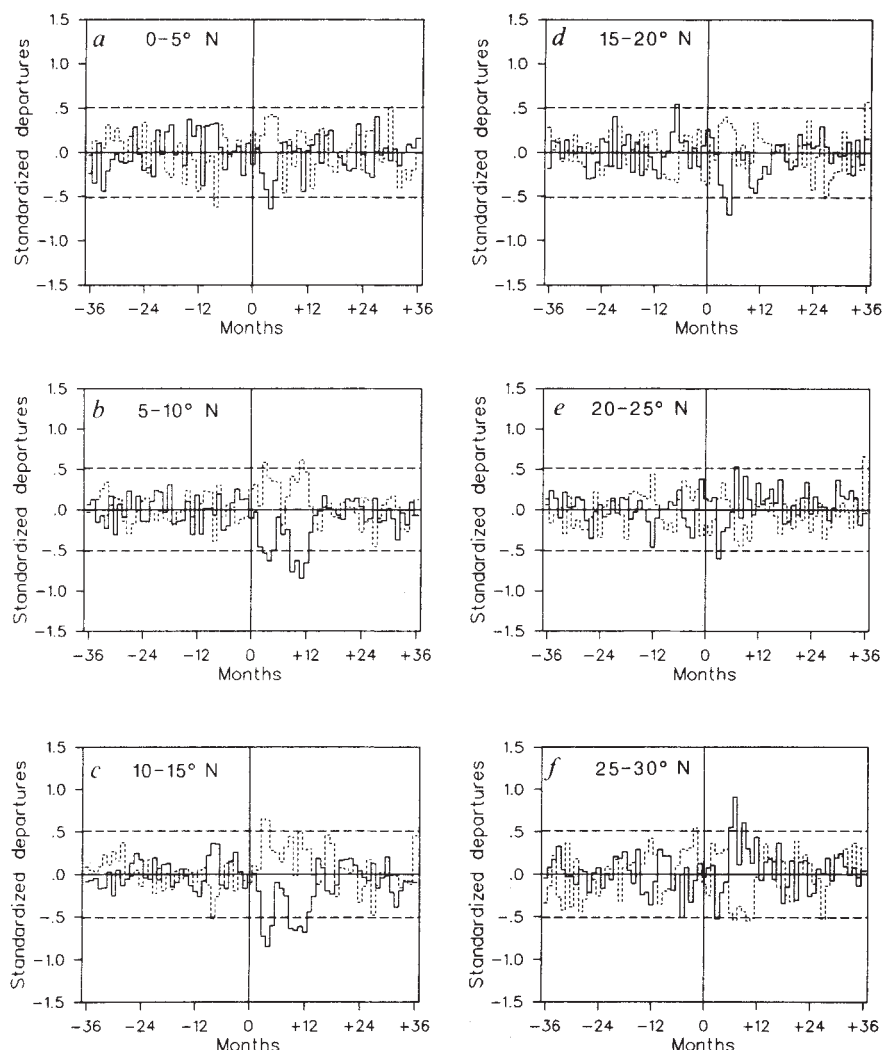


Fig. 5 As in Fig. 3, except for zonal precipitation anomalies. Precipitation anomalies (from 1921 to 1960 mean) normalised by standard deviation of respective month for period 1881-1980.

is strongest and occurs latest in the Asian sector (Fig. 4c), while in the American sector, the signal is stronger during cold rather than warm events (Fig. 4). Thus, the signals of Figs 3 and 4 are not due to any particular sector, but appear to occur globally.

The superposed epoch analysis was repeated to determine the time variation of large-scale precipitation anomalies associated with ENSO events (Figs 4d-f and 5). During warm events, precipitation over the zonally averaged continental areas of the Northern Hemisphere tropics tends to be low, with two distinct anomalies (Fig. 5). The first occurs during the northern summer monsoon season, and reflects the tendency for below normal precipitation to occur over southern Asia^{25-29,31,32} and the Sahel region of Africa³⁰. This is clearly illustrated in the regional graphs (Fig. 4d-f). The source of the second minimum around March of year +1 is primarily from the tropical Asian sector (Fig. 4f), although the signal is also evident in the other two sectors. The changes during cold events are generally the reverse of those during warm events. The positive (negative) precipitation anomaly north of 20° N occurs in the winter months during year +1 of warm (cold) events. It is due principally to an increase (decrease) in precipitation over the southeastern United States (Fig. 4e) during warm (cold) events. Anomalies over the African continent exhibit the least coherence and signal strength of the three regions.

As these regional data show, there is considerable variability across longitudes not only with regard to the amplitude of the anomalies, but also, in some instances with regard to the sign⁵. Nevertheless, it is clear that ENSO events have a powerful influence on large-scale temperature and precipitation patterns

in the Northern Hemisphere. ENSO-related anomalies in the Southern Hemisphere are also significant and have been detected far from the original source of forcing^{5,10}.

The large spatial scale and long-lived nature of the surface temperature and precipitation anomalies, together with the fact that the climatic response has the opposite tendency during ENSO extremes make these findings of some practical importance. The swings in temperature and precipitation associated with ENSO contribute a large fraction of the interannual variability in the tropics and portions of the extratropics.

This work was funded by the Carbon Dioxide Research Division, US Department of Energy.

Received 31 October 1986; accepted 23 March 1987.

1. Troup, A. J. *Q. Jl R. met. Soc.* **91**, 490-506 (1965).
2. Bjerknes, J. *Month. Weath. Rev.* **97**, 163-172 (1969).
3. Horel, J. D. & Wallace, J. M. *Month. Weath. Rev.* **109**, 813-829 (1981).
4. van Loon, H. & Rogers, J. C. *Month. Weath. Rev.* **109**, 1163-1168 (1981).
5. van Loon, H. & Madden, R. A. *Month. Weath. Rev.* **109**, 1150-1162 (1981).
6. Rasmusson, E. M. & Carpenter, T. H. *Month. Weath. Rev.* **109**, 1150-1162 (1982).
7. Yarnal, B. *Progr. Phys. Geogr.* **9**, 315-352 (1985).
8. Kiladis, G. N. & Diaz, H. F. *Month. Weath. Rev.* **114**, 1035-1047 (1986).
9. Yarnal, B. & Diaz, H. F. *J. Climat.* **6**, 197-219 (1986).
10. van Loon, H. & Shea, D. J. *Month. Weath. Rev.* **113**, 2063-2074 (1985).
11. Philander, S. G. H. *J. Atmos. Sci.* **42**, 2652-2662 (1985).
12. Bradley, R. S., Kelly, P. M., Jones, P. D., Goodess, C. M. & Diaz, H. F. *A Climatic Data Bank for Northern Hemisphere Land Areas 1851-1980*, Tech. Rep. 017 (US Dept. of Energy, Washington, DC, 1985).
13. Jones, P. D. *et al. J. Clim. appl. Met.* **25**, 161-179 (1986).
14. Bradley, R. S. *et al. Science* (in the press).
15. Chen, W. Y. *Mon. Weath. Rev.* **110**, 800-807 (1982).
16. Parker, D. E. *Met. Mag.* **112**, 184-188 (1983).
17. Jones, P. D. & Ropelewski, C. F. *Month. Weath. Rev.* (in the press).
18. Newell, R. E. & Weare, B. C. *Nature* **262**, 40-41 (1976).

19. Angell, J. K. *Month. Weath. Rev.* **109**, 230–243 (1981).
20. Angell, J. K. & Korshover, J. *Month. Weath. Rev.* **111**, 901–921 (1983).
21. Pan, Y. H. & Oort, A. H. *Month. Weath. Rev.* **111**, 1244–1258 (1983).
22. Parker, D. E. *J. Climat.* **5**, 273–282 (1985).
23. Conrad, V. & Pollock, L. W. *Methods in Climatology* (Harvard University Press, Cambridge, Massachusetts, 1950).
24. Fu, C., Diaz, H. F. & Fletcher, J. O. *Month. Weath. Rev.* **114**, 1716–1738 (1986).
25. Quinn, W. H., Zopf, D. O., Short, K. S. & Kuo Yang, R. T. W. *Fish. Bull.* **76**, 663–678 (1978).
26. Wright, P. B. *An Index of the Southern Oscillation* (Climatic Research Unit Publication CRU RP4, Univ. of East Anglia, Norwich, UK, 1975).
27. Rasmusson, E. M. & Wallace, J. M. *Science* **222**, 1195–1202 (1983).
28. Meehl, G. A. *Month. Weath. Rev.* (in the press).
29. Berlage, H. P. *Mededelingen en Verhandelingen* No. 69 (Koninklijk Netherlands Meteorologische Instituut, Staatsdrukkerij, S'Gravenhage, Netherlands, 1957).
30. Ropelewski, C. F. & Halpert, M. S. *Month. Weath. Rev.* (in the press).
31. Bhalme, H. N. & Jadhav, S. K. *J. Climat.* **4**, 509–520 (1984).
32. Parthasarathy, B. & Part, G. B. *J. Climat.* **5**, 369–378 (1985).

Kilometre-scale thermohaline overturn of pore waters in the Louisiana Gulf Coast

Jeffrey S. Hanor

Department of Geology, Louisiana State University, Baton Rouge, Louisiana 70803, USA

Convincing evidence exists for the large-scale convection of sea water through the upper oceanic crust, even at considerable distances from mid-ocean ridges^{1,2}. There has been considerable interest in verifying the existence of convective overturn of pore fluid in large sedimentary basins, a mass-transport process which could profoundly influence the transport of heat, hydrocarbons, metals and diagenetically reactive dissolved constituents^{3,4}. The existence of fluid convection in sediments of the United States Gulf Coast has been inferred on the basis of the large volumes of fluid necessary to account for the observed diagenetic alteration⁵ and the existence of thermal anomalies that cannot be explained by simple conductive transport of heat^{3,6}. Theoretical Rayleigh–Darcy calculations^{6,7} support the notion that thermally driven convection is possible. Here I present geochemical and physical evidence for the existence of density inversions in Gulf Coast pore fluids sufficient to drive large-scale convective fluid flow at rates conceivably as high as metres per year. The density inversions are caused in part by the dissolution of salt diapirs and the formation of dense, saline brines at shallow depths.

The Iberia salt dome in Iberia Parish, Louisiana, is one of many piercement salt diapirs in the Louisiana region of the United States Gulf Coast sedimentary basin. The dome extends to within 290 m of the ground surface, and the upper 2.5 km of salt penetrates a middle Miocene to Pleistocene series of hydro-pressured sands and shales⁸. Hydrocarbon production extends over a depth interval exceeding 2 km, and chemical analyses of brines co-produced with these hydrocarbons provide a unique cross-sectional look at variations in pore fluid properties in this portion of the Gulf Coast.

Among the dissolved constituents of interest in terms of the subsurface flow regime at Iberia are the monofunctional volatile fatty acids (VFAs), which include acetic, propionic, butyric and valeric acids and their dissociated acid anions and complexes. The highest concentrations of VFAs in basinal waters are found at burial depths corresponding to temperatures of 80 to 120 °C⁹ (below depths of 2.5 km at Iberia), where the VFAs are produced by thermal degradation of organic matter. At shallower depths and lower temperatures, VFAs are destroyed by decarboxylation and bacterial fermentation^{9,10}. The spatial distribution of VFAs in pore waters at Iberia^{8,11} (Fig. 1) supports the hypothesis that a tongue of VFA-rich pore fluid is migrating upwards along the flank of the salt dome from a source at depth. The concentration of VFAs along this flow path decreases as thermal and bacterial degradation proceed.

The isotopic composition of the pore fluids at Iberia¹² is also consistent with the hypothesis of large-scale vertical mixing. The spatial variation in $\delta^{18}\text{O}$ (Fig. 1) can most easily be explained by the upward migration along the flank of the dome of diagenetically altered waters enriched in heavy oxygen and the mixing of these waters with a descending mass of fluid having a pronounced meteoric oxygen signature. There are slight deflections in the isotherms consistent with advective heat transport along the inferred flow paths (Fig. 1). The thermal picture is complicated, however, by the contrast in thermal conductivity of the salt diapir and the surrounding sediments.

Fluid overturn at Iberia appears to be driven by density inversions resulting from increasing temperature with depth and spatial variations in salinity (Fig. 1). Preferential dissolution of salt is occurring near the top of the dome. Less saline water is bleeding up into the section from the geopressured zone below⁸. Fluid densities at Iberia have been calculated for *in situ* conditions using known relations (S. L. Phillips *et al.* Lawrence Berkeley Lab. Report LBL-12810, 1981) between the density of NaCl solutions and pressure, temperature and salinity. Fluid pressures were calculated assuming hydrostatic equilibrium with a fluid of constant density of 1.04 g cm³. Corrected bottom-hole temperature measurements are limited to the control points shown in Fig. 1. Linear temperature gradients were assumed to exist between the ground surface and the top of control. Direct salinity measurements are also limited to the control points in Fig. 1. Additional indirect salinity determinations were made from spontaneous potential (SP) logs at approximately 100–250 m intervals. The vertical lines in Fig. 3 show the extent of log control. The salinity structure of the upper 300 m is not known, but was assumed to be similar to known regional variations¹³.

The inferred density field for NaCl solutions in west-east cross-section A–A' is shown in Fig. 2. Of greatest significance is the fact that densities progressively decrease below a depth of 600 to 1,000 m in this section, chiefly as a result of increasing temperature. The whole water column below these depths is gravitationally unstable, and the hydraulic potential thus exists for driving convective overturn.

If the pore fluid pressure and density fields are known then it is possible to deduce flow paths and the magnitude of the hydraulic gradient driving fluid flow. A general expression for Darcy's law is

$$\mathbf{V} = -(k/\phi\eta)(\nabla P + \rho g \nabla z) \quad (1)$$

where vector $\mathbf{V}$ is the velocity and direction of fluid flow, k is the intrinsic sediment permeability, ϕ is sediment porosity, η is dynamic fluid viscosity, P is fluid pressure, ρ is fluid density, g is the gravitational constant, and z is elevation¹⁴. Following Hubbert¹⁵ we define the hydraulic force field, $\mathbf{H}$ (pressure/length or force/volume) as

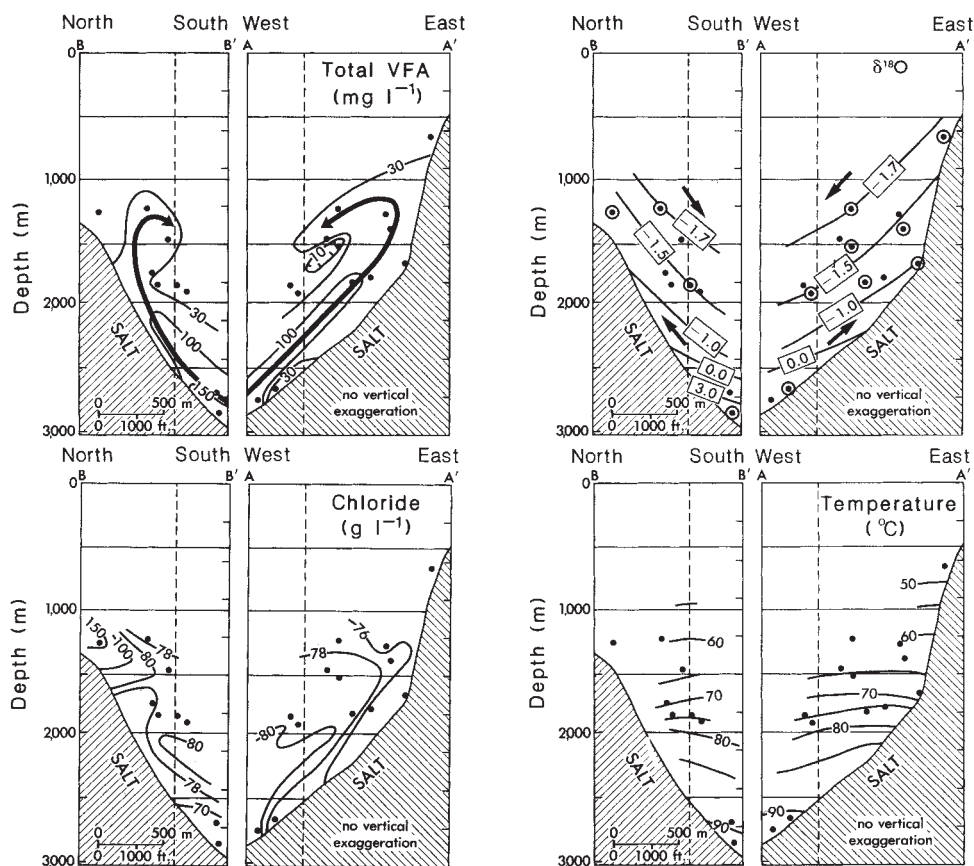
$$\mathbf{H} = -(\nabla P + \rho g \nabla z) \quad (2)$$

Note that the force field, $\mathbf{H}$, is not the hydraulic head gradient ∇h (length/length), a parameter used to describe isodensity flow systems¹⁴. The density field shown in Fig. 3 defines the term $(\rho g \nabla z)$. In the absence of direct pressure measurements, an approximate configuration of the pressure field can be obtained by assuming that all fluids are in hydrostatic equilibrium, so

$$P_z = P(\text{hydrostatic}) \equiv P_0 - g \int_0^z \rho(z) dz \quad (3)$$

The fluids at Iberia cannot truly be in hydrostatic equilibrium, of course, because a hydrostatic system as defined by equation (2) will not convect $(\partial P(\text{hydrostatic})/\partial z = -\rho g \nabla z)$. However, the pressure field obtained by this simple approximation allows an order of magnitude estimate to be made of the horizontal component of the hydraulic force field, $(\partial H/\partial x)$, which in turn sets a lower limit on the magnitude of $\mathbf{H}$.

Fig. 1 Spatial variation in total dissolved volatile fatty acids (VFAs)⁸, $\delta^{18}\text{O}$ (ref. 12), dissolved chloride, and corrected¹³ bottom-hole temperatures of aqueous pore fluids on the south-west flank of the Iberia salt dome (Fig. 4). West-east cross-section A-A' and north-south section B-B' intersect along the vertical dashed line. Circled black dots show control points on the $\delta^{18}\text{O}$ diagram, plain dots show control on the other diagrams. Arrows show the inferred direction of fluid migration. The arrows are schematic; the actual flow path is more complex and tortuous.



For the purposes of numerical analysis, the two-dimensional study cross-section A-A' was divided into a finite-element mesh (Fig. 2). Density values at the nodes of each element of this mesh, as determined from Fig. 2, were used to calculate P from equation (3). Refinement of calculated density *in situ* could have been made using these new pressure values, but was deemed unnecessary. Discrete values of P and ρ at these nodes were used to calculate $(\partial H/\partial x)$ (equation (2)) for the central points of each element using conventional numerical techniques. Results are shown in Fig. 3. According to this simple approximation, there are three fluid domains: a no-flow region at the top, an intermediate domain where the horizontal component of flow is to the west (left in the diagram), and a lower domain where the horizontal component of flow is to the east (to the right in the diagram). These two domains correspond to the downwelling and upwelling portions of the convective system as inferred from geochemistry (Fig. 1). Maximum calculated values of $(\partial H/\partial x)$ are in the range of 50–60 Pa m^{-1} . If the pressure field were known more accurately, the calculated vectors shown in Fig. 3 would probably have vertical components.

Note that the calculated position of the convergent boundary between the two lower domains coincides spatially with a region of high shale content which separates two thick masses of sand-dominated sediment. Based on preliminary evaluation of spatial variations in pore fluid properties over the entire flank of the dome, it is probable that section A-A' actually transects a more complex, three-dimensional convective system pictured schematically in Fig. 4. Although there is evidence for the existence of one or two nearly vertical faults of small displacement in section A-A', there is no evidence that they are important in controlling flow direction. In contrast, faults at some other salt domes currently under investigation¹⁶ appear to act as preferential channelways for flow. Critics of the convective hypothesis¹⁷ argue that it is not possible to drive large volumes of fluid at high velocity across shale beds. I agree. Flow at Iberia

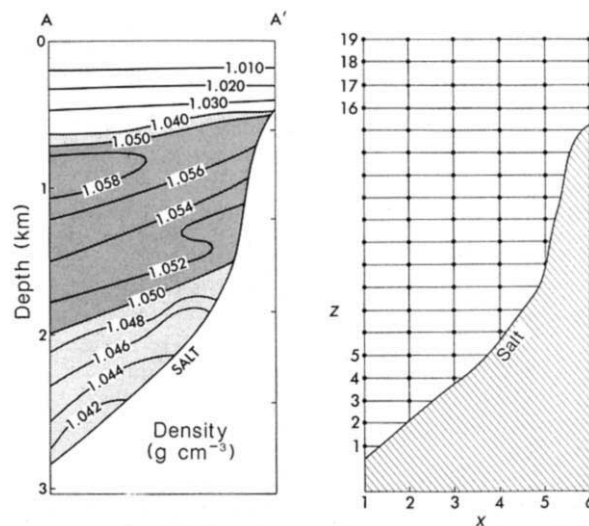


Fig. 2 Spatial variation in calculated pore fluid density in section A-A' (left) and finite element grid used to calculate hydraulic gradients (right). Densities were calculated (S. L. Phillips *et al.* Lawrence Berkeley Lab Report LBL-12810, 1981, ref. 13) for *in situ* temperatures, pressures, and NaCl-equivalent salinities using information derived and interpolated from spontaneous potential/resistivity logs and from direct analysis. Densities in the upper 300 m of the section inferred from a previous regional study¹³

is undoubtedly taking place along a highly complex and tortuous network of interconnected sands (Fig. 3), not along the schematic cross-formational arrows shown in Fig. 1.

It is possible to make some preliminary order of magnitude estimates of the rate of fluid flow from equation (1). A hydraulic gradient of 50 Pa m^{-1} (Fig. 3), an intrinsic permeability of

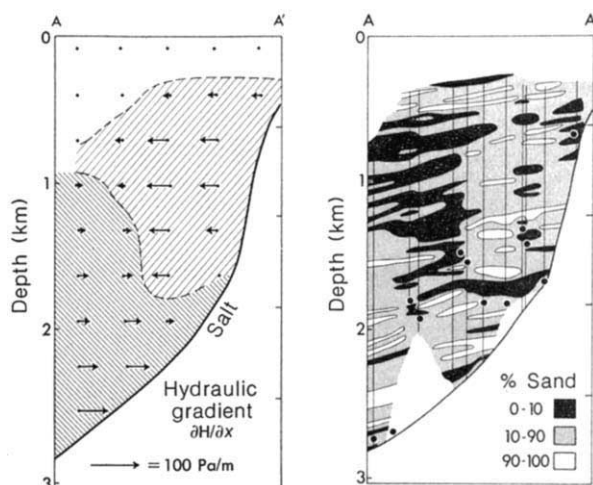


Fig. 3 Variation in the horizontal component of the hydraulic force field, H , in the plane of section A-A' (left) and the spatial variation in the distribution of sands and shales (right), as inferred from spontaneous potential response. The arrows in the hydraulic gradient diagram are orientated in the direction of presumed fluid flow. The lowest domain corresponds to the region of upwelling pictured in Fig. 1, and the middle domain corresponds to the region of downwelling. The convergent boundary between these two domains corresponds spatially to a sand-poor, shale-rich zone of presumed low permeability (black areas).

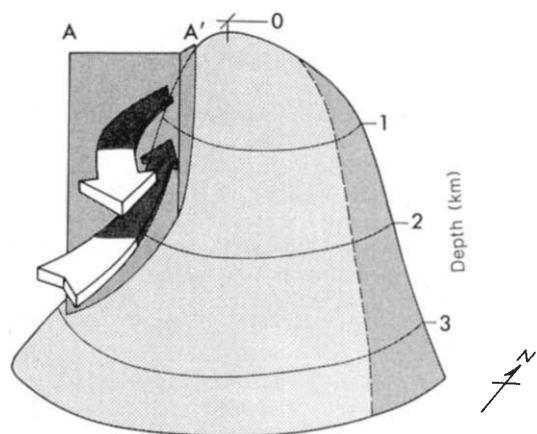


Fig. 4 Three-dimensional sketch of the Iberia salt dome, no vertical exaggeration, showing the location of section A-A' and the possible gross spatial configuration of the upwelling and downwelling limbs of the flow system documented in Figs 1-3. The schematic flow paths are inferred from an evaluation of three dimensional variations in pore fluid chemistry.

10^{-12} m^2 (equivalent to 1 darcy), a porosity of 0.2 and a viscosity of $6 \times 10^{-4} \text{ N s m}^{-2}$ yield from equation (1) a flow rate of 13 m per yr. Although a permeability of one darcy may seem high, it is typical of many shallow sands in this portion of the Gulf Coast¹⁸ and at Iberia specifically (R. Gerdes, personal communication). The actual flow field at Iberia is undoubtedly more complex and spatially variable than that represented by these initial calculations. With the acquisition of more detailed field information on temperatures, porosities, permeabilities, and initial shut-in fluid pressures, it will be possible to model the fluid flow and diagenetic regimes at Iberia in a more rigorous, three-dimensional fashion. The present simple calculations, however, support the geochemical evidence for large-scale convective overturn of fluids at Iberia.

Thermohaline circulation is theoretically possible in any

sedimentary basin where shallow dissolution of salt is occurring. The density inversion documented in Fig. 2 is part of a density inversion of regional extent¹³, and it is thus possible that vertical overturn is likewise of regional importance in the south Louisiana Gulf Coast sedimentary basin. If salt dissolution is restricted to the base of a sedimentary sequence, however, as in the northern Gulf Coast¹⁹, there may be no density inversion and thus no fluid overturn.

I thank Texaco Inc. for permission to sample fluids at Iberia. I also thank A. Workman and H. Posey for comments on chemical aspects of this study, V. Ranganathan, D. Evans and M. Simms for discussions of the physics, and R. Gerdes for information on permeabilities at Iberia. M. Rodgers helped with data reduction.

Received 8 January; accepted 3 April 1987.

1. Anderson, R. N., Langseth, M. G. & Sclater, J. C. *J. geophys. Res.* **82**, 3391-3409 (1977).
2. Williams, C. F., Naramsimhan, T. N., Anderson, R. N., Zoback, M. D. & Becker, K. *J. geophys. Res.* **91**, 4877-4899 (1986).
3. Aziz, K., Bories, S. A. & Combarnous, M. A. *J. Can. Petrol. Technol.* **12**, 41-47 (1973).
4. Cassan, J. P., Garcia Palacios, M., Fritz, B. & Tardy, Y. *Bull. Cent. Rech. Elf-Aquit.* **5**, 113-135 (1981).
5. Land, L. S. *Mem. Am. Ass. Petrol. Geol.* **37**, 47-62 (1984).
6. Blanchard, P. E. & Sharp, J. M. *Trans. S.W. Sect. Am. Ass. Petrol. Geol.* **6-12** (1985).
7. Wood, J. R. & T. A. *Geochim. cosmochim. Acta* **46**, 1707-1713 (1982).
8. Workman, A. L. & Hanor, J. S. *Trans. Gulf Coast Ass. geol. Soc.* **35**, 293-300 (1985).
9. Carothers, W. W. & Kharaka, Y. K. *Bull. Am. Ass. Petrol. Geol.* **62**, 2441-2453 (1978).
10. Surdam, R. C., Boese, S. & Crossey, L. J. *Mem. Am. Ass. Petrol. Geol.* **37**, 127-149 (1984).
11. Hanor, J. S. & Workman, A. L. *Appl. Geochem.* **1**, 37-46 (1986).
12. Posey, H. H., Workman, A. L., Hanor, J. S. & Hurst, S. D. *Trans. Gulf Coast Ass. geol. Soc.* **35**, 261-268 (1985).
13. Hanor, J. S., Bailey, J. E., Rogers, M. C. & Milner, L. R. *Trans. Gulf Coast Ass. geol. Soc.* **36**, 143-149 (1986).
14. Garven, G. & Freeze, R. A. *Am. J. Sci.* **284**, 1085-1124 (1984).
15. Hubbert, M. K. *Trans. Am. Inst. Min. Metall. Petrol. Engrs* **207**, 222-239 (1956).
16. Bennett, S. & Hanor, J. S. in *Dynamical Geology of Salt and Related Structures* (eds Lerche, I. & O'Brien, J. J.) 425-440 (Academic, New York, 1987).
17. Bjorlykke, K. *Mem. Am. Ass. Petrol. Geol.* **37**, 277-288 (1984).
18. Bebout, D. G. & Gutierrez, D. R. *Proc. U.S. Gulf Coast Geopress.-Geotherm. Energy Conf.* **5**, 13-24 (eds Bebout, D. G. & Bachman, A., Louisiana State Univ., Baton Rouge, 1981).
19. Hanor, J. S. *Trans. Gulf Coast Ass. geol. Soc.* **34**, 55-61 (1984).

CO₂, melts and granulite metamorphism

B. Ronald Frost & Carol D. Frost

Department of Geology and Geophysics, University of Wyoming,
Laramie, Wyoming 82071, USA

The discovery that granulites contain CO₂-rich fluid inclusions while amphibolites in the same terrane contain H₂O-rich or mixed CO₂-H₂O inclusions¹ has led to a debate over the processes that produce granulite metamorphism. One theory²⁻⁴ maintains that granulites form as the result of a massive influx of CO₂ from lower crustal or mantle sources. Such an influx could be responsible for the marked depletion many granulite terranes show in large-ion lithophile elements^{2,5}. Others, however, maintain that granulite-facies metamorphism is a necessary complement of crustal melting⁶⁻⁹. Recently, Lamb and Valley^{10,11} have argued that granulite metamorphism may be a product of any one of three processes: CO₂-streaming, partial melting, or recrystallization of originally dry rocks. Here we suggest that these processes need not be independent and that they all may be related to the passage of melts through the crust.

Numerous experiments show that CO₂ has a finite, albeit limited, solubility in silicate melts (Table 1). This solubility is dependent on the melt composition and decreases with decreasing pressures and temperatures¹²⁻¹⁹. Those melts which are less strongly polymerized will dissolve greater amounts of CO₂ because the CO₂ reacts with the non-bridging oxygens to form a carbonate ion^{14,20,21}. Because H₂O tends to de-polymerize melts, CO₂ is more soluble in hydrous melts than in anhydrous melts of the same composition^{14,18}. Since CO₂ is much less

soluble in silicate melts than H_2O , melts that are saturated in CO_2 will crystallize at higher temperatures than those that are H_2O -saturated. Furthermore, since the CO_2 solubility of melts is greater at high pressure and temperature^{14,15,22}, melts emplaced into the crust will become saturated with a CO_2 -rich vapour phase long before H_2O saturation is attained^{13,23}. If the fluid is progressively removed from the magma body as the melt rises, then the magma will be able to remain molten down to lower temperatures and the exsolved fluid will become progressively enriched in H_2O . Along with CO_2 , the early fluids will also be enriched in other components, such as NaCl , that are not strongly soluble in the silicate melt¹³. If sufficient NaCl were present in the fluid, it could lead to the unmixing of the fluid to a CO_2 -rich vapour and a saline brine even at temperatures in excess of 1,000 °C (refs 24, 25). This is important because such unmixing could further enrich the CO_2 content of the vapour phase which, being less dense than the associated brine, will be more likely to infiltrate and dehydrate the country rocks.

Fluid inclusion studies of charnockitic granitoids provide evidence for CO_2 -saturated melts. They show that the fluid composition changes from carbonic to aqueous as the rock evolves from charnockite to biotite granite²⁶. In fact, in some of the granitoids studied, the presence of a two-phase fluid is suggested by the occurrence of both carbonic and saline fluid inclusions²⁷⁻³⁰. This evidence that charnockites can evolve CO_2 -rich vapours which dehydrate the country rock is corroborated by the occurrence of narrow zones of granulite metamorphism around thin charnockite dikes in the Wind River Mountains of Wyoming³¹ (Fig. 1) and in South India³².

The arguments above suggest that a substantial amount of CO_2 may move through the Earth's crust associated with melts. We contend that the CO_2 postulated to have streamed through some terranes²⁻⁴ could have originated from CO_2 -bearing melts emplaced into the crust. The basic aspects of this model are shown in Fig. 2. The figure depicts a rifting environment of relatively high heat flow (see also ref. 33), although granulites are almost certainly also associated with the deeper portions of magmatic arcs, as, for example, in the Sierra Nevada³⁴. Basaltic magma from the mantle rises into the crust. Some portions of this magma move through the crust to be erupted on the surface. Although they may not be CO_2 -saturated initially, during their passage the magmas will become CO_2 -saturated either through decompression and cooling or through crystallization. On saturation, they will evolve a CO_2 -rich fluid phase that will move into the adjacent country rocks. In addition, heat from this magma may melt the portions of the country rock more prone

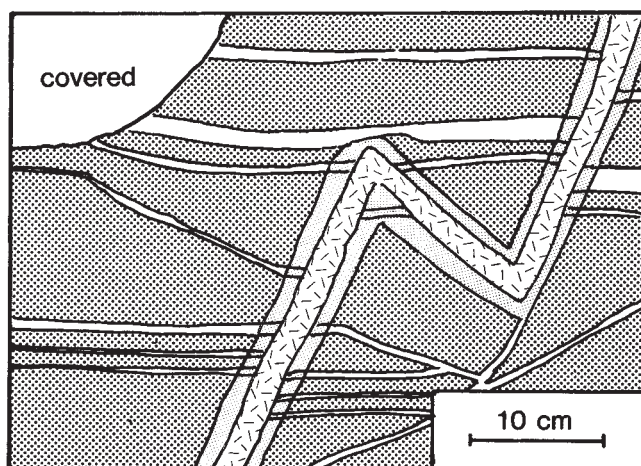


Fig. 1 Field sketch showing late charnockite dikelet (dashes) producing granulite (light stippling) in amphibolite (heavy stippling), Medina Mountain area, Wind River Mountains, Wyoming. Blank areas are feldspathic layers in amphibolite. Amount of amphibolite dehydrated by this dike indicates that the fracture must have been traversed by several volumes of melt or that it was a conduit that carried both melt and CO_2 -rich magmatic fluids.

to fusion with the resulting relatively hydrous melt being incorporated into the basaltic magma³⁵. The end result will be dehydration in the vicinity of the feeder. A large portion of the basaltic magma, however, is likely to pond in the lower crust, where it will produce melting of the adjacent crustal material. Because the temperatures involved lie well above the H_2O -saturated granite minimum, crustal melts produced in this environment are likely to be vapour-undersaturated or will coexist with fluids that are CO_2 -rich. As with the mafic magmas, movement of these granitic melts through the crust will tend to produce CO_2 -rich fluids while H_2O is absorbed by the melt, either in the form of a crustal melt or directly as a fluid. Granitic melts emplaced into granulite environments cannot crystallize to the H_2O -saturated solidus because this lies below the ambient temperature. Any crystallization that takes place will leave a more hydrous melt that must move to higher crustal levels before solidifying³⁶. The lower portions of batholiths formed from this melt may be hot enough to crystallize as charnockite, but at upper levels the melt will crystallize to form hornblende- or biotite-bearing granites. In such a model, the granitic melt

Table 1 Mass balance for CO_2 and granulites

A, Global carbon flux and potential for dehydration of crust

Present-day carbon flux from the mantle⁴⁰

Estimated present-day carbon flux from the mantle through the continents^{40,41} *

Volumes of rock dehydrated

Potential volume of crust dehydrated

$1-8 \times 10^{12}$ moles per yr

$1-8 \times 10^{11}$ moles per yr

640 cm^3 biotite granite per mol CO_2

130 cm^3 amphibolite per mol CO_2

0.5 km^3 biotite granite per yr

0.1 km^3 amphibolite per yr

B, Granulite/magma volume ratios

Solubility of CO_2 in magma (wt%)

Moles CO_2 per km^3 magma

Volume ratio mafic granulite/magma

Volume ratio felsic granulite/magma

Basaltic magma
($\rho = 2.7$)

Granitic magma
($\rho = 2.4$)

1.7% (30 kbar, 1450 °C)¹⁵

0.5% (3 kbar, 1200 °C)¹²

9×10^{11}

3×10^{11}

0.1

0.04

0.6

0.2

* Assumes that the proportion of mantle carbon exhaled through the continents is similar to the proportion of ^3He vented through the continents. The carbon flux may have been greater in the past than at the present⁴⁰.

† Calculations use equation (4) of Greenwood⁴² and assume a porosity of 0.1% and that biotite granite has 5% biotite and amphibolite has 50% amphibole. CO_2 -streaming is assumed to have changed the ambient fluid from $X(\text{CO}_2) = 0$ to $X(\text{CO}_2) = 0.8$ (ref. 43) with complete dehydration of silicates.

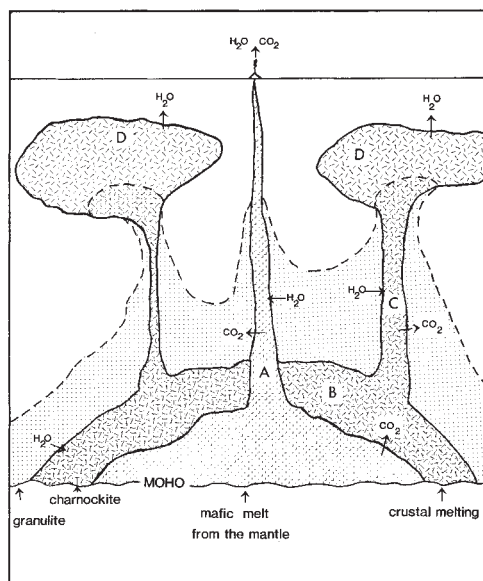


Fig. 2 Diagram showing the role of melts in dehydration of the lower crust. A, Basaltic magma is emplaced into the crust. As it ascends it evolves a CO_2 -rich fluid and incorporates H_2O from the country rock. B, Basaltic magma ponded at the base of the crust will induce melting in the country rock. C, Felsic melts thus produced will rise into the crust, evolving CO_2 and absorbing H_2O . Portions of these magmas that crystallize at depth will form charnockites or other pyroxene-bearing rocks. D, Felsic magmas ascending to shallow levels will balloon out to form batholiths, which will be dominated by hydrous granitoids.

becomes both a conduit for moving CO_2 into the lower crust and a means for moving H_2O to higher crustal levels.

Although CO_2 -streaming may not be the volumetrically dominant process of granulite formation, the validity of this model depends on whether significant volumes of granulite can be produced from the amounts of CO_2 known to dissolve in silicate melts. The estimates in Table 1 are minima because the calculations address only dehydration by passive CO_2 -streaming and do not consider the dehydrating effects of the partial melts that are likely to be produced during CO_2 -streaming³⁷. If the carbon flux through the continents interacts with the crust on ascent, 0.1–0.5 km^3 granulite can be produced each year (Table 1). Like the carbon flux through the ocean ridges, the carbon flux through the continents and any associated granulite formation will be concentrated along zones of active magmatism. If basaltic melts are saturated with CO_2 at 30 kbar, only 1 km^3 of melt can contain the total annual continental CO_2 flux. Likewise, 1 km^3 of superheated granitic melt at 3 kbar can accommodate a large proportion of this CO_2 flux (Table 1). These results suggest that melts in the mantle and lower crust can accommodate all of the CO_2 observed to be expelled through the crust.

Whereas previous theories on the origin of granulites have been disparate or even contradictory, the model presented here involves one simple process that can accommodate all modes of granulite formation. Not only will magmas provide a source for CO_2 , they may also crystallize to form large volumes of pyroxene-bearing igneous rocks. The presence of these dry magmas has been documented in the Wind River Mountains³⁸, and we contend that they may be present in other areas where later deformation has eliminated the igneous textures of rocks crystallized from these melts.

This model implies that granulite metamorphism is directly tied to an igneous event, a conclusion also suggested by the pressure and temperature conditions recorded by granulite mineral assemblages (Fig. 3). The pressure-temperature conditions estimated for granulite terranes require elevated geothermal gradients. Extensional regimes, such as the Basin and Range,

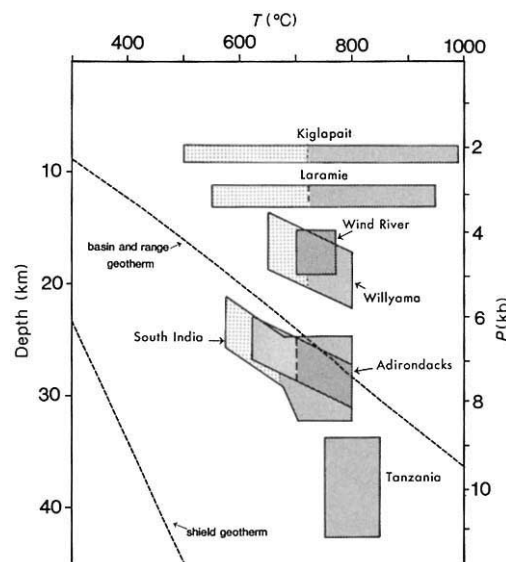


Fig. 3 Depth versus temperature for some granulite and amphibolite terranes compared to Shield and Basin and Range geotherms⁴⁴. Heavy symbols, granulite facies; light symbols, amphibolite facies. Transition is shown in dashed lines if temperature is well known, dotted line if it is poorly constrained. Data from refs 45 (Kiglapait), 46 (Laramie), 31 (Wind River), 47 (Willyama), 39 (Adirondacks), 48 (S. India), 49 (Tanzania).

have higher geothermal gradients that encompass some granulite conditions, but this does not necessarily imply that all granulites formed in such regimes. They could equally well be produced in areas of lower regional geothermal gradient if the geotherms have been locally elevated by the passage of magma.

There is a complete gradation from low-pressure contact-metamorphism environments to high-pressure granulite terranes (Fig. 3). In the low-pressure terranes the participation of a melt as a heat source is easily recognized, both by the high temperatures encountered and the isobaric nature of the transition from amphibole-bearing to pyroxene-bearing assemblages. In a slightly higher-pressure terrane, the Wind River Mountains, the role of melts as a source of fluids is evident (Fig. 1)^{31,38} and even in the relatively high-pressure regime of the Adirondacks a magmatic heat source is postulated³⁹.

In conclusion, granulites formed by CO_2 -streaming, melting or metamorphism of syn-metamorphic dry igneous rocks may represent different facets of a single process involving passage of melts through the crust. As each sub-process may produce a characteristic geochemical signature, we must constrain the geochemical changes accompanying each of these sub-processes before we can understand the geochemical effects of granulite metamorphism as a whole.

This research was supported by the NSF. We would also like to thank John Valley, Bob Newton, Steve Wickham and Jim Myers whose critical reviews of this manuscript greatly improved the presentation of our theories.

Received 22 December 1986; accepted 23 March 1987.

1. Touret, J. *Lithos* **4**, 423–436 (1971).
2. Heier, K. S. *Phil. Trans. R. Soc. A273*, 429–442 (1973).
3. Collerson, K. D. & Fryer, B. S. *Contr. Miner. Petrol.* **67**, 151–167 (1978).
4. Newton, R. C. & Hansen, E. C. *Mem. geol. Soc. Am.* **161**, 167–178 (1983).
5. Tarney, J. & Windley, B. F. *J. geol. Soc.* **134**, 153–172 (1977).
6. Brown, G. C. & Fyfe, W. S. *Contr. Miner. Petrol.* **240**, 310–318 (1970).
7. Brown, G. C. & Fyfe, W. S. *Int. Geol. Cong.* **24–2**, 27–34 (1972).
8. Fyfe, W. S. *Phil. Trans. R. Soc. A273*, 457–461 (1973).
9. Nesbitt, H. W. *Contr. Miner. Petrol.* **72**, 303–310 (1980).
10. Lamb, W. & Valley, J. *Nature* **312**, 56–58 (1984).
11. Lamb, W. & Valley, J. in *The Deep Proterozoic Crust in the North Atlantic Provinces* (eds Tobin, A. C. & Touret, J. L. R.) 119–131 (Reidel, Dordrecht, 1985).
12. Kadik, A. A., Lukanin, O. A., Lebedev, Ye. B. & Korovushkina, E. Y. *Geochem. Int.* **9**, 1041–1050 (1972).
13. Kadik, A. A. & Lukanin, O. A. *Geochem. Int.* **10**, 115–129 (1973).

14. Mysen, B. O. *Am. J. Sci.* **276**, 969–996 (1976).
15. Mysen, B. O., Arculus, R. J. & Eggler, D. H. *Contr. Miner. Petrol.* **53**, 227–239 (1975).
16. Mysen, B. O., Eggler, D. H., Seitz, M. G. & Holloway, J. R. *Am. J. Sci.* **276**, 455–479 (1976).
17. Brey, G. *Contr. Miner. Petrol.* **57**, 215–221 (1976).
18. Eggler, D. H. & Rosenhauer, M. *Am. J. Sci.* **278**, 64–94 (1978).
19. Eggler, D. H. & Kadik, A. A. *Am. Miner.* **64**, 1036–1048 (1979).
20. Fine, G. & Stolper, E. *Contr. Miner. Petrol.* **91**, 105–121 (1985).
21. Fine, G. & Stolper, E. *Earth planet. Sci. Lett.* **76**, 263–278 (1986).
22. Spera, F. J. & Bergman, S. C. *Contr. Miner. Petrol.* **74**, 55–66 (1980).
23. Holloway, J. R. *Geol. Soc. Am. Bull.* **87**, 1513–1518 (1976).
24. Bowers, T. S. & Helgeson, H. C. *Geochim. cosmochim. Acta* **47**, 1247–1275 (1983).
25. Skippen, G. & Trommsdorff, V. *Am. J. Sci.* **286**, 81–104 (1986).
26. Konnerup-Madsen, J. *Am. J. Sci.* **277**, 637–696 (1977).
27. Falkum, T., Konnerup-Madsen, J. & Rose-Hansen, J. in *The Deep Proterozoic Crust in the North Atlantic Province* (eds Tobi, A. C. & Touret, J. L. R.) 579–583 (NATO ASI, Ser. C, No. 158, 1985).
28. Konnerup-Madsen, J. *Bull. Miner.* **107**, 327–340 (1984).
29. Konnerup-Madsen, J. & Rose-Hansen, J. *Chem. geol.* **37**, 79–93 (1982).
30. Konnerup-Madsen, J., Debussy, J. & Rose-Hansen, J. *Lithos* **18**, 271–280 (1985).
31. Koesterer, M. E. *et al. Precamb. Res.* (in the press).
32. Ravindra Kumar, G. R. & Chacko, T. *J. geol. Soc. India* **28**, 277–288 (1986).
33. Sandiford, M. & Powell, R. *Earth planet. Sci. Lett.* **79**, 151–158 (1986).
34. Ross, D. C. *Geology* **13**, 288–291 (1985).
35. Waters, D. J. & Whales, C. J. *Contr. Miner. Petrol.* **88**, 269–275 (1984).
36. Field, D., Drury, S. A. & Cooper, D. C. *Lithos* **13**, 281–289 (1980).
37. Grant, J. A. *Am. Miner.* **71**, 1071–1075 (1986).
38. Hulsebosch, T. P., Koesterer, M. E. & Frost, B. R. *Geol. Soc. Am. Abstr. Progr.* **17**, 616 (1985).
39. Bohlen, S. R., Valley, J. W. & Essene, E. J. *J. Petrology* **26**, 971–992 (1985).
40. Des Marais, D. J. in *Atmospheric CO₂: Natural Variations, Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 602–611 (Am. Geophys. Un. Monogr. Ser. **32**, Washington DC, 1985).
41. O'Nions, R. K. & Oxburgh, E. R. *Nature* **306**, 429–431 (1983).
42. Greenwood, H. J. *Am. J. Sci.* **273**, 561–571 (1975).
43. Wells, P. R. A. *J. Petrology* **20**, 187–226 (1979).
44. Sclater, J. G., Jaupart, C. & Galson, D. *Rev. geophys.* **18**, 269–311 (1980).
45. Berg, J. H. & Docka, J. A. *Am. J. Sci.* **283**, 414–434 (1983).
46. Grant, J. A. & Frost, B. R. *Geol. Soc. Am. Abstr. Progr.* **18**, 620 (1986).
47. Phillips, G. N. & Wall, V. J. *Bull. Miner.* **104**, 801–810 (1981).
48. Raase, P., Raith, M., Ackermann, D. & Lal, R. K. *J. Geol.* **94**, 261–282 (1986).
49. Newton, R. C. *Am. J. Sci.* **283-A**, 1–28 (1983).

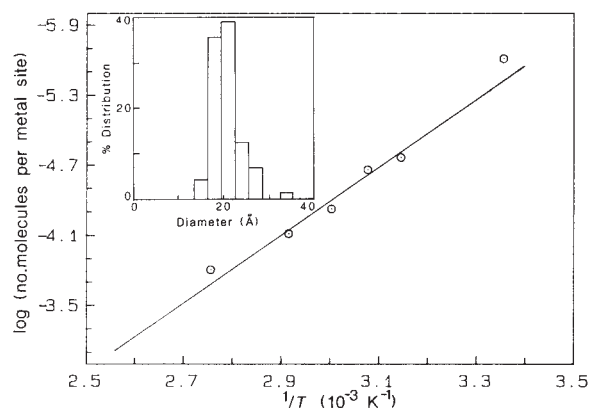


Fig. 1 Effect of temperature on the rate of dark methanation of CO₂. Conditions: $p_{\text{CO}_2} = 0.05$ atm, $p_{\text{H}_2} = 0.6$ atm, $p_{\text{total}} = 1$ atm, total volume 20 ml. Turnover frequencies determined at the beginning of the reaction are plotted logarithmically against reciprocal absolute temperature. Inset: histogram showing particle size distribution of the catalyst, consisting of ~75% RuO_x ($x \leq 2$) and 25% Ru as shown by XPS using a Physical Electronics Perkin-Elmer model 550/590 spectrometer, MG-K α line, 1,253.6 eV, width 0.7 eV. Ru-loaded TiO₂ powder was mounted on a 6 × 6 mm-sized In-foil and kept under Ar during the transfer to the spectrometer; temperature of the sample during the analysis was ~298 K and pressure $< 10^{-7}$ torr; elements detected were: Ti, Ru, In, C and O; Ru 3d-5/2 line was used to discriminate between the metallic and higher valency states of Ru. XPS spectrum displayed two features at 281.0 and 279.8 eV which can be assigned to Ru(III or IV) and Ru(O), respectively; ratio of oxidized to metallic Ru was estimated from the relative intensities of these bands.

Methanation and photo-methanation of carbon dioxide at room temperature and atmospheric pressure

K. Ravindranathan Thampi, John Kiwi & Michael Grätzel*

Institut de Chimie Physique, Ecole Polytechnique Fédérale, CH-1015 Lausanne, Switzerland

The Sabatier reaction:



is an important catalytic process of wide industrial and academic interest^{1–3}. It is applied to syngas conversion and the treatment of waste streams. Methane is one of the most important carbon resources of the world, serving as an energy vector as well as a feedstock for higher-value chemicals^{4–6}. Despite its favourable thermodynamics, the eight-electron reduction of CO₂ to CH₄ by hydrogen is difficult to achieve: high-energy intermediates impose large kinetic barriers, and the formation of side products is common. Intensive investigations during the past decade have therefore been aimed at improving the activity and selectivity of methanation catalysts^{1–3,8–11}. Although significant progress has been made in this field, elevated temperatures ($T > 300^\circ\text{C}$) and pressures ($P > 1$ atm) are still required for methane generation to proceed at significant rates and yields. Here we report the selective conversion of CO₂ to methane at room temperature and atmospheric pressure, using highly dispersed Ru/RuO_x loaded onto TiO₂ as a catalyst. The reaction rate is sharply enhanced through photo-excitation of the support material.

Experiments were conducted with TiO₂-P25 powder, a mixture of ~80% anatase and 20% rutile (BET surface area 55 m² g⁻¹, density 3.8 g cm⁻³), which was obtained from Degussa, West Germany. Loading with Ru (3.8% confirmed by

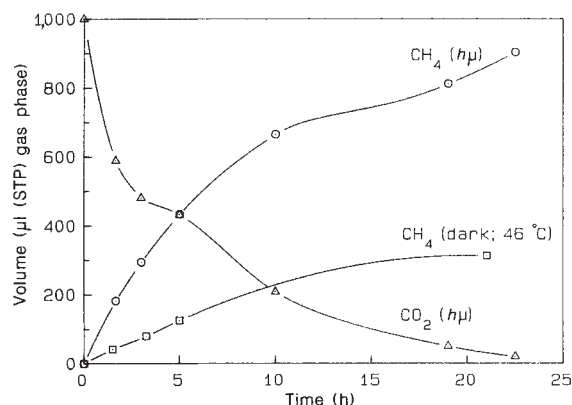
elemental analysis) was carried out as follows: 100 mg RuCl₃·3H₂O (alpha inorganic; Ru content 38%) was dissolved in 100 ml 0.1 M HCl and the solution left to equilibrate for 1 day. TiO₂ (1 g) was added and the pH increased to 4.5 by means of a 0.1 M KOH solution. The temperature of the dispersion was raised to 60 °C and maintained over a period of 5 h during which the pH was repeatedly readjusted to 4.5. Water was subsequently evaporated by heating the dispersion in an open dish until it boiled and the solid was calcined for 20 h at 170 °C and 12–18 h at 375 °C. Residual KCl was removed from the catalyst by dialysis which was carried out over a period of two days. Subsequently, the sample was dried overnight at 110 °C. This procedure yields RuO₂ particles of 10–20 Å size¹¹. The RuO₂ was partially reduced in a H₂/Ar (1:1) stream (flow rate 40 ml h⁻¹) at 220 °C for 1 h yielding a catalyst consisting of ~75% RuO_x ($x \leq 2$) and 25% Ru, as shown by X-ray photoelectron spectroscopy (XPS) analysis. Reduction of RuO₂ to Ru is complete if carried out at 500 °C. However, the activity of such a catalyst as well as unreduced RuO₂ in the methanation reaction was found to be inferior to that consisting of the Ru/RuO_x mixture.

The dispersion of Ru determined by conventional H₂ adsorption measurement was 50–55%, assuming a 1:1 stoichiometry for the chemisorption of hydrogen atoms by surface Ru. This high dispersion value excludes strong metal-support interactions which suppress H₂ adsorption¹². This is expected from the relatively low temperature employed in the reduction which renders the formation of such a state very unlikely. Transition electron microscopy (Phillips EM 400) revealed the presence of catalyst particles with an average diameter of 20 Å. A representative histogram for the size distribution function is given in the inset of Fig. 1. Assuming that the particles have a spherical shape, the calculated dispersion is 65% which is in fair agreement with the experimental results¹³.

Methanation experiments employed 100 mg catalyst powder which was spread over the bottom of a flat Pyrex vessel (volume

* To whom correspondence should be addressed.

Fig. 2 Photo-methanation of CO_2 at 46°C in the solar simulator. The decrease in CO_2 and increase in CH_4 volume (STP) is plotted as a function of illumination time. For comparison, the dark evolution of CH_4 is shown. The conditions are the same as in Fig. 1.



20 cm^3) equipped with a side arm and a septum for the admission and withdrawal of gas samples. Exposure of the catalyst to air after reduction was avoided by always maintaining it under argon. The transfer to the reactor was performed in a glove box, and reaction temperature was controlled by a microprocessor.

Methane formation was first examined in the dark. The reactor initially contained 1 ml CO_2 ($p_{\text{CO}_2} = 0.05\text{ atm}$) and 12 ml H_2 ($p_{\text{H}_2} = 0.6\text{ atm}$). The remaining gas was argon with traces of nitrogen, the total pressure being 1 atm. Surprisingly, methane was formed in the presence of the TiO_2/Ru catalyst at a temperature as low as 25°C . By contrast, SiO_2/Ru and $\text{Al}_2\text{O}_3/\text{Ru}$ prepared in the same manner and tested under identical conditions as TiO_2/Ru were found to be inactive for room temperature methanation. Methane formation and CO_2 consumption obeyed strictly the 1:1 stoichiometry predicted by equation (1) indicating that the catalyst operates in a very selective fashion.

This important and surprising finding was substantiated by employing highly sensitive and specific analytical techniques, that is, gas chromatography, mass spectrometry and high-pressure liquid chromatography (HPLC), which failed to detect other by-products. In particular, the formation of carbon monoxide, methanol, formaldehyde, ethane and higher homologues can be excluded within the detection limit which for these compounds was at least $0.002\text{ }\mu\text{mol}$ per μmol of methane generated. A special effort was made to check whether formic acid or oxalic acid were produced. When methanation was complete any water soluble product was extracted from the reaction system with 1 ml water. After the catalyst was centrifuged the supernatant was analysed by HPLC. Blank experiments showed the detection limit to be a factor of 500 below the quantity of CH_4 produced in a typical run. Thus, the overall selectivity of the catalyst is better than 99%. Initial rates of methane formation were $0.17\text{ }\mu\text{mol h}^{-1}$ at 25°C and $10.5\text{ }\mu\text{mol h}^{-1}$ at 90°C . These rates correspond to turnover frequencies (TF) of $2.5 \times 10^{-6}\text{ s}^{-1}$ and $1.6 \times 10^{-4}\text{ s}^{-1}$, respectively, based on 50% Ru dispersion. An Arrhenius plot of the TF values measured in the temperature range from 25 – 90°C is given in Fig. 1. From the slope of the straight line the activation energy (E_A) is determined as 54 kJ mol^{-1} . Published E_A -values for the methanation of CO_2 range from 30 to 171 kJ mol^{-1} , depending on catalyst and support material^{1–3,7–10,14}. All the earlier studies were conducted at elevated temperature ($T \geq 400\text{ K}$). The room temperature conversion of CO_2 to methane has so far not been observed.

TiO_2 is a semiconductor (bandgap 3 eV) which absorbs near-ultraviolet radiation. This opens up the possibility of affecting the methanation reaction (1) by light excitation of the support material. To our knowledge this effect has not been explored before. Photo-methanation experiments employed the same conditions as in Fig. 1 except that the TiO_2/Ru catalyst was subjected to band gap illumination. Figure 2 shows results obtained with a solar simulator (total intensity 0.08 W cm^{-2}) as the light source.

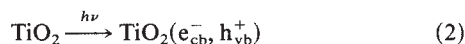
The temperature of the catalyst measured with a sensor was 46°C . Methane was produced at an initial rate of $116\text{ }\mu\text{l h}^{-1}$ (STP). Within 5 h of photolysis, $450\text{ }\mu\text{l}$ of CH_4 had been produced and the same amount of CO_2 was left in the gas phase. (The observation in Fig. 2 that during photo-methanation the number of moles of methane and CO_2 present in the gas phase add up to less than the number of moles of CO_2 initially injected is due to co-adsorption of CO_2 and hydrogen on the surface of the catalyst. This was verified by measuring the adsorption isotherms of CO_2 in the absence and presence of hydrogen.) Dark methanation at 46°C occurred at rates 4–5 times less than the initial rate. The thermal methanation levels off after one day when $\sim 30\%$ of the CO_2 had reacted. The decrease in the rate is attributed to the accumulation of liquid water which appears to retard the reaction. By contrast, the photo-conversion of CO_2 to methane was almost complete after 22 h. At this point, the gas mixture was flushed out from the reactor and H_2O removed in an Ar stream maintained at 100°C . After injection of 1 ml CO_2 and 12 ml H_2 the illumination in the solar simulator was repeated. The generation of methane occurred at the same rate as during the first run. By continuing this procedure, a total volume of 11 ml ($\approx 0.5\text{ mmol}$) of CH_4 , corresponding to a turnover number of 28 with respect to the surface Ru atoms, was produced without decline in the catalytic activity.

The stoichiometry of the Sabatier reaction was preserved ($<99\%$) under photo-excitation of the support, and careful analysis of the reaction products provided no evidence for the formation of compounds other than methane as stated before. With $p_{\text{CO}_2} = 1\text{ atm}$ and temperature at 46°C the extrapolated turnover frequency was $1.4 \times 10^{-3}\text{ s}^{-1}$. This is a strikingly high TF value for the methanation of CO_2 under such mild conditions (low temperature and ambient pressure).

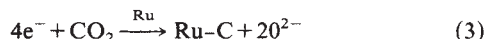
To elucidate the mechanism of this important photo effect, we investigated which part of the light spectrum was active in enhancing the rate of methanation. These experiments were carried out at 25°C using a 150-W high pressure Xe lamp equipped with a water jacket to remove infrared radiation. When a 695 nm cutoff filter was placed in the light beam, methanation proceeded at essentially the same rate as in the dark, at $0.17\text{ }\mu\text{mol h}^{-1}$. Changing to a 435 nm cutoff filter increased the rate only slightly to $0.27\text{ }\mu\text{mol h}^{-1}$. However, the rate tripled to $0.54\text{ }\mu\text{mol h}^{-1}$ when no filter was used and bandgap radiation admitted to the sample. In this case, the short ultraviolet radiation ($\lambda < 310\text{ nm}$) is removed by the Pyrex wall of the reactor. Since the fraction of light emitted by the lamp in the 310–435 nm region is less than 5% of the total intensity, these experiments show that bandgap excitation of the TiO_2 support makes the dominant contribution to the photo enhancement of the methanation reaction. Therefore, redox processes involving electron-hole pairs are likely to be important in the photo-reaction. These may lead to the generation of high-energy intermediates which are difficult to form in the dark at room temperature. Photo

effects induced by bandgap excitation of semiconductor supports have been observed in catalytic reactions¹⁵⁻¹⁸, and this field has been reviewed¹⁹.

A key role in the methanation of CO₂ on Ru is played by carbidic surface carbon (Ru-C). The enthalpy of Ru-C formation is 100 kJ mol⁻¹ positive with respect to graphite and does not depend on surface coverage²⁰. This indicates the existence of isolated carbon atoms coordinated to Ru without the formation of graphitic islands or overlayers. Photogeneration of such Ru-C species would involve bandgap excitation of the TiO₂ resulting in electron-hole pair formation:



followed by reduction of CO₂



The concomitant valence band process could involve hydrogen oxidation



followed by neutralization



Methane is subsequently generated in the reaction of surface carbon with hydrogen:



The thermal methanation could proceed by a similar mechanism. But the generation of carbidic surface carbon would involve the thermal reduction of adsorbed CO₂ by coadsorbed hydrogen:



The mechanism postulated here is plausible in view of the fact that reactions (3) and (6) have been found to occur in the cathodic reduction of CO₂ to methane on Ru electrodes²¹ and colloids (I. Willner, personal communication). However, photo-electronic effects such as the increase in the electrical conductivity of the support under light excitation could also be important

This work was supported by the Gas Research Institute, Chicago, Illinois (subcontract with the Solar Energy Research Institute, Golden, Colorado). We thank Drs A. McEvoy and H. Mathieu for their valuable assistance in the ESCA experiments, and Dr P. Cuendet for his help in the HPLC analysis. We also thank Mr Amanz Ruf for help with the gas chromatography and mass spectrometry analysis. M.G. acknowledges discussions with Professor G. Ertl, Fritz-Haber-Institute, Berlin.

Received 5 December 1986; accepted 3 April 1987.

- Lunde, P. J. & Kester, F. L. *J. Catal.* **30**, 423-429 (1973).
- Phyng Quack, T. Q. & Rouleau, D. *J. appl. Chem. Biotechnol.* **26**, 527-535 (1976).
- Tomsett, A. D., Hagiwara, T., Miyamoto, A. & Inui, T. *Appl. Catal.* **26**, 391-394 (1986).
- Keim, W. in *Chemistry of the Future* (ed. Grunewald, H.) 53-62 (Pergamon, Oxford, 1983).
- Tomoyasu, I. & Lunsford, J. H. *Nature* **314**, 721-722 (1985).
- Otsuka, K., Jinno, K. & Morikawa, A. *J. Catal.* **100**, 353-359 (1986).
- Solymosi, F., Erdöheli, A. & Kocsis, M. *JCS Faraday Trans. 1* **77**, 1003-1012 (1981).
- Solymosi, F., Erdöheli, A. & Bausagi, T. *J. Catal.* **68**, 371-382 (1981).
- Weatherbee, G. D. & Bartholomew, C. H. *J. Catal.* **87**, 352-362 (1984).
- Inui, T., Funabiki, F., Suehiro, M. & Sezume, T. *JCS Faraday Trans. 1* **75**, 787-802 (1979).
- Blondeel, G., Harriman, A., Porter, G., Urwin, D. & Kiwi, J. *J. phys. Chem.* **87**, 2629-2636 (1983).
- Tauster, S. J., Fung, S. C. & Garten, R. L. *J. Am. chem. Soc.* **100**, 170-175 (1978).
- Anderson, J. R. *Structure of Metallic Catalysts* 361 (Academic, London, 1973).
- Gupta, N. M., Kamble, V. S. & Iyer, R. M. *J. Catal.* **66**, 101-111 (1980).
- Yixuan, C. *et al. J. molec. Catal.* **21**, 275-280 (1983).
- Kawai, T. & Sakata, T. *Nature* **282**, 283-284 (1979).
- Cunningham, J., Tobin, P. J. & Meriaudeau, P. *Surface Sci.* **108**, L465-L469 (1981).
- Vijayakumar, K. M. & Lichten, N. N. *J. Catal.* **90**, 173-177 (1984).
- Cunningham, J. in *Comprehensive Chemical Kinetics* Vol. 19, Ch. 3 (eds Bamford, C. H., Tipper, C. F. H. & Compton, R.) (Elsevier, Amsterdam, 1984).
- Wise, H. & McCarthy, J. G. *Surface Sci.* **133**, 311-320 (1983).
- Frese, K. W. & Leach, S. *J. electrochem. Soc.* **132**, 259-260 (1985).

Effect of exothermicity on electron transfer rates in photosynthetic molecular models

A. D. Joran, B. A. Leland, P. M. Felker, A. H. Zewail, J. J. Hopfield & P. B. Dervan

Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA

To appreciate fully the significance of the high-resolution crystal structure of the reaction-centre complex from the purple bacterium *Rhodospseudomonas viridis*^{1,2}, there is increasing impetus to construct models to ascertain how the specific variables such as distance³⁻⁵, exothermicity⁶⁻⁹ and solvent¹⁰⁻¹¹ control the rates of electron transfer through the integral membrane protein complex. Recently, we have determined the effect of a 4 Å change in distance (at fixed exothermicity, solvent and temperature) on the rates of photon-induced electron transfer for two porphyrin-quinone assemblies separated by rigid spacers^{3,4}. From picosecond fluorescence lifetime measurements on zinc *meso*-phenyloctamethylporphyrin coupled to a benzoquinone moiety via one and two bicyclo[2.2.2]octyl spacers (10 and 14 Å, edge-to-edge), we determined that the rates of electron transfer are 10¹⁰ s⁻¹ and ≤ 10⁷ s⁻¹, respectively (the exothermicity, -ΔG = 1.0 eV in acetonitrile at 25 °C). This result revealed that the porphyrin-quinone separated by a 10-Å spacer would be ideally suited, with regard to picosecond fluorescence techniques for a study of exothermicity effects at fixed distance and solvent. Here we report the fluorescence lifetimes of a homologous series of seven porphyrin-quinone molecules, each with differently substituted benzoquinones separated by an identical rigid phenylbicyclo[2.2.2]octane spacer (10 Å, edge-to-edge) which vary with respect to driving force for electron transfer from the first excited singlet state of the porphyrin. The key features of this series of molecules lie in the fact that the edge-to-edge distance is fixed and ΔG is tunable. Thus, the effect of exothermicity on electron transfer can be measured while avoiding major perturbations of the electronic structure.

Porphyrin-quinones 1-7 (Fig. 1) were prepared by a general route⁴ involving the condensation of a tetrapyrrolic *ac*-biladiene dihydrobromide¹² with different dihydroxy- or dimethoxyaryl-bicyclo[2.2.2]octylbenzaldehyde derivatives⁴, followed by deprotection of the methyl ethers (BI₃), oxidation (PbO₂), and metallation with zinc acetate. In addition, seven 2-methyl-1,4-benzoquinones substituted in the 5- and 6-positions (Table 1) were synthesized for use as models in electrochemical studies. Full synthetic details will be reported elsewhere. All structures were characterized by ultraviolet-visible, ¹H nuclear magnetic resonance (NMR) and mass spectral data.

The half-wave potentials for reduction of the seven model benzoquinones were determined in acetonitrile (Table 1). From this series of substituted quinones and the corresponding electrochemical data in acetonitrile for the [5-(4'-*tert*-butylphenyl)-2,3,7,8,12,13,17,18-octamethylporphyrinato]zinc(II), the porphyrin-quinones 1-7 were ranked with regard to exothermicity for electron transfer from the first excited singlet state of each compound. Because these data are not corrected for coulomb interactions or ion-pairing effects, the exothermicity ranking is correct in a relative order but not with regard to absolute scale, particularly in non-polar solvents where these values may be large. In each solvent series an exothermicity range of 0.6 eV for photon-induced electron transfer can be examined.

Fluorescence lifetimes for solutions containing porphyrins 1-7 in acetonitrile, butyronitrile, 2-methyltetrahydrofuran and benzene were determined at 298 K with a mode-locked cavity-

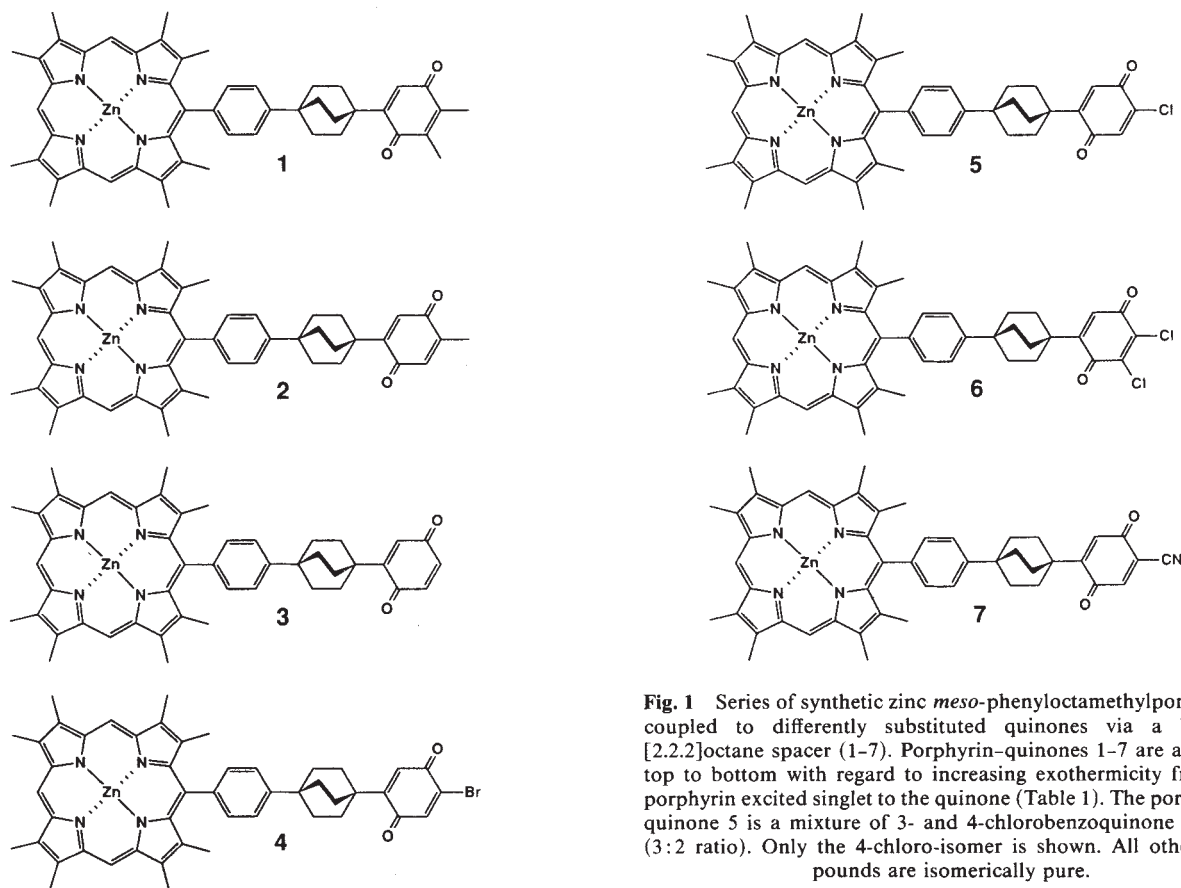


Fig. 1 Series of synthetic zinc *meso*-phenyloctamethylporphyrins coupled to differently substituted quinones via a bicyclo[2.2.2]octane spacer (1-7). Porphyrin-quinones 1-7 are arranged top to bottom with regard to increasing exothermicity from the porphyrin excited singlet to the quinone (Table 1). The porphyrin-quinone 5 is a mixture of 3- and 4-chlorobenzoquinone isomers (3:2 ratio). Only the 4-chloro-isomer is shown. All other compounds are isomerically pure.

dumped picosecond dye laser (rhodamine 6G) system. The picosecond apparatus used in these studies has been described elsewhere¹³. The laser pulses were typically 15 ps in duration and the full output power was ~30 mW. The fluorescence emission was detected by a fast photomultiplier using time-correlated single photon counting techniques. The quality of the data fits to single or double exponentials (see below) was excellent as judged by the χ^2 criterion.

Chromatography was used just before the picosecond fluorescence lifetime experiments and checked by NMR for purity.

Table 1 Half-wave potentials for reduction of 2-methyl-1,4-benzoquinone substituted by 5- R_1 , 6- R_2 groups in acetonitrile

— R_1 , — R_2		$E_{1/2}(Q/Q^{\cdot-})(V)^*$
—Me, —Me		–1.12
—Me, —H		–1.04
—H, —H		–0.95†
—Br, —H		–0.79
—Cl, —H		–0.78
—Cl, —Cl		–0.66
—CN, —H		–0.55

Potentials were determined by linear sweep voltammetry with 25 or 127 μ diameter Pt microelectrode using 0.1 M tetrabutylammonium perchlorate (TBAP) as supporting electrolyte, Pt wire auxiliary electrode, and ferrocene/ferrocenium reference electrode.

* Half-wave potentials are certain to within ± 15 mV and are reversible or nearly reversible [$E_{3/4} - E_{1/4}$] = 55–75 mV ($n = 1$).

† Determined by differential pulse voltammetry (ref. 3).

The absorption spectra of porphyrin-quinones 1-7 in solution are well described by the sum of the individual chromophores and are insensitive to the addition of the spacer while the wavelengths of the emission spectra are that of the unperturbed porphyrin. Deaerated solutions of porphyrins 1-7 at $\sim 10^{-5}$ – 10^{-6} M concentrations were excited at 570 nm and the emission was observed at 580 and/or 630 nm. The observed decays can be fit to biexponentials with one major fast component (55–829 ps) assigned to the porphyrin-quinone and a minor component with longer lifetime (~ 1.5 ns) assigned to the porphyrin-hydroquinone^{3,4}.

From the time-resolved fluorescence decays, the major fast exponential (τ_1), was attributed to the deactivation of the porphyrin-excited state by intramolecular electron transfer to the quinone. In all four solvents there was a decrease in the fluorescence lifetime followed by an increase with increasing exothermicity (Table 2). Rate constants for electron transfer, calculated from $k_{et} = 1/\tau_1 - 1/\tau_0$, where τ_0 is the observed fluorescence lifetime of the porphyrin in the absence of electron transfer, show a trend of sharply increasing rates with increasing driving force from –0.8 to –1.0 eV (acetonitrile, uncorrected for coulombic effects). Over the next 0.3 eV the rates of electron transfer are similar and finally at 1.4 eV the rate of electron transfer decreases (Fig. 2).

The Marcus expression¹⁴ of the form

$$k = A \exp \left[\frac{-(\Delta G + \lambda)^2}{4\lambda k_B T} \right] \quad (1)$$

where^{15,16}

$$A = \frac{2\pi}{\hbar} |T_{ab}|^2 \left(\frac{1}{\sqrt{4\pi\lambda k_B T}} \right) \quad (2)$$

$$\lambda = \lambda_i + \lambda_o \quad (3)$$

was used for a least-squares analysis of the data (Fig. 2). The

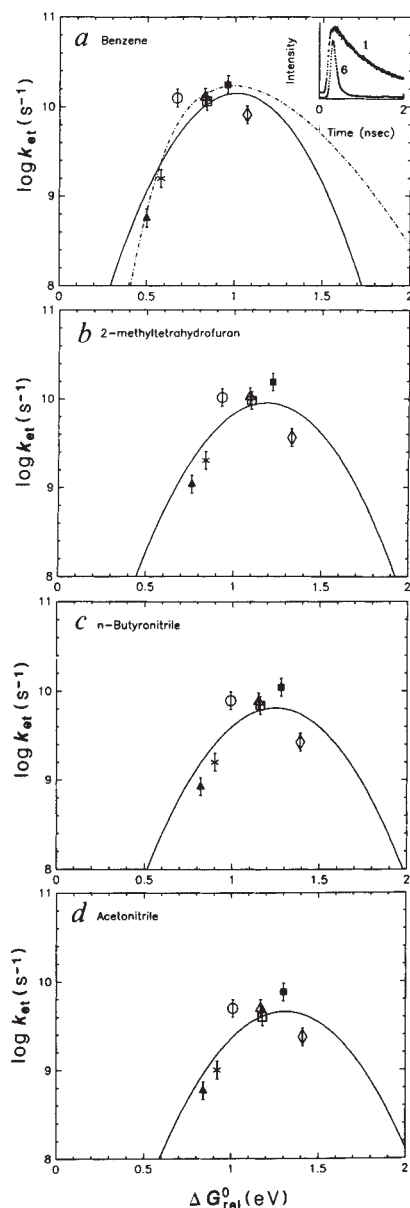


Fig. 2 Plot of $\log(k_{\text{et}})$ against $-\Delta G^0$ (eV) (Table 2). Solid line is a least squares fit of Marcus theory (see equation (1)). Fit parameters (T_{ab} , λ): a, benzene (7.4 cm^{-1} , 1.01 eV); b, 2-methyltetrahydrofuran (6.1 cm^{-1} , 1.19 eV); c, butyronitrile (5.3 cm^{-1} , 1.25 eV); d, acetonitrile (4.5 cm^{-1} , 1.31 eV). Dashed line (Fig. 2a) is semi-classical theory (equation (4), see text for parameters). Inset in upper right corner (Fig. 2a) is a comparison of the picosecond fluorescence decays of 1 and 6.

data for each solvent were fitted by two free parameters, the tunnelling matrix element governing the electron transfer (T_{ab}) and the reorganization energy (λ). The values obtained in the four solvents have the following ranges: $T_{\text{ab}} = 5$ to 7 cm^{-1} and $\lambda = 1.0$ to 1.3 eV (see legend to Fig. 2). The λ values obtained are large by ~ 0.5 eV for porphyrin-quinone transfers¹⁶⁻²⁰. This may reflect the fact that the absolute values of ΔG shown in Fig. 2 are uncorrected for the coulomb interactions in the charge separated state and the ion-pairing effects in the electrochemical measurements. The tunnelling matrix element (T_{ab}) governing the transfer is similar to a value of $\sim 10 \text{ cm}^{-1}$ calculated for the initial charge separation event in reaction centres for transfer from the bacteriochlorophyll dimer to the pheophytin acceptor¹⁹. Qualitatively, the data appear to follow the expected trend of an increase in λ with increasing solvent polarity. The very weak dependence of the electron transfer rate for a given com-

Table 2 Fluorescence lifetimes for compounds 1-7

Cmpd*/Solvent†	τ (ns)‡	k_{et} (s^{-1})§	$-\Delta G_{\text{ref}}^0$ (eV)
1a	0.801, 1.722(9:1)	5.68×10^8	0.50
2a	0.443, 1.161(20:1)	1.58×10^9	0.58
3a	0.076, 1.490(71:1)	1.25×10^{10}	0.67
4a	0.075, 1.362(50:1)	1.27×10^{10}	0.83
5a	0.082, 1.549(200:1)	1.15×10^{10}	0.84
6a	0.055, 1.894(167:1)	1.75×10^{10}	0.96
7a	0.114, 1.520(1:2)	8.09×10^9	1.07
1b	0.594, 1.488(7:1)	1.08×10^9	0.76
2b	0.384, 1.591(16:1)	2.00×10^9	0.84
3b	0.092, 1.537(11:1)	1.03×10^{10}	0.93
4b	0.090, 1.611(20:1)	1.05×10^{10}	1.09
5b	0.099, 1.661(77:1)	9.50×10^9	1.10
6b	0.062, 1.629(42:1)	1.55×10^{10}	1.22
7b	0.236, 1.717(2:3)	3.63×10^9	1.33
1c	0.679, 1.798(4:1)	8.44×10^8	0.82
2c	0.448, 1.295(7:1)	1.60×10^9	0.90
3c	0.118, 1.467(20:1)	7.85×10^9	0.99
4c	0.121, 1.512(8:1)	7.64×10^9	1.15
5c	0.132, 1.576(25:1)	6.95×10^9	1.16
6c	0.085, 1.445(23:1)	1.11×10^{10}	1.28
7c	0.302, 1.811(1:1)	2.68×10^9	1.39
1d	0.829, 1.569(2:1)	5.89×10^8	0.84
2d	0.617, 1.430(6:1)	1.01×10^9	0.92
3d	0.178, 1.601(20:1)	5.00×10^9	1.01
4d	0.178, 1.513(5:1)	5.00×10^9	1.17
5d	0.215, 1.491(2:1)	4.03×10^9	1.18
6d	0.121, 1.283(24:1)	7.65×10^9	1.30
7d	0.335, 1.660(1:1)	2.37×10^9	1.41

* All compounds absorb at $\lambda_{0-0} = 571 \pm 1$ nm in benzene. The first two emission bands occur at 578 and 631 nm. In other solvents, the absorption peak shifts slightly: 2-methyltetrahydrofuran (575 nm); butyronitrile (574 nm); acetonitrile (574 nm). Fluorescence was filtered and detected at right angles at 580 and 630 nm.

† Solvent a, benzene; b, 2-methyltetrahydrofuran; c, butyronitrile; d, acetonitrile. Benzene and acetonitrile were HPLC grade and distilled from CaH_2 . 2-Methyltetrahydrofuran was distilled from benzophenone ketyl. Butyronitrile was distilled from $\text{KMnO}_4/\text{Na}_2\text{CO}_3$ and then from P_2O_5 .

‡ Biexponential lifetime data listed as τ_1 , τ_2 , and amplitude fraction (1:2). In all cases, the quality of the fits to biexponential decays were excellent as judged by the χ^2 criterion (< 1.15) for individual transients. In Fig. 2 we present the larger estimate of the reproducibility error ($\pm 10\%$ of the lifetime) and not the standard deviation.

§ Calculated from $k_{\text{et}} = (1/\tau_{\text{obsd}}) - (1/\tau_0)$ where τ_0 is the natural fluorescence lifetime of the porphyrin and is equal to 1.47, 1.59, 1.66, and 1.64 ns for C_6H_6 , $n\text{PrCN}$, MTHF, and CH_3CN , respectively (see also ref. 3).

|| Calculated from $-\Delta G^0 = -E_{1/2}(\text{P}/\text{P}^{++}) + E_{1/2}(\text{Q}/\text{Q}^{--}) + W^*$ where $W^* = 2.15 \pm 0.01$ eV is the singlet excitation energy, $E_{1/2}(\text{P}/\text{P}^{++})$ from reference 3, $E_{1/2}(\text{Q}/\text{Q}^{--})$ from Table 1 and ref. 3 and is uncorrected for coulomb interactions in the charge-separated state.

pound with changes in solvent polarity (for example, a factor of ~ 2 for 6a against 6d in Table 2) is expected for the generation of a charge-separated state via electron transfer from an initially neutral species²¹. Conversely, dramatic polarity effects are expected to be observed in the back transfer rates in this series.

A semi-classical model in which the solvent is treated classically, the donor is assumed to be in the vibrational ground state, and quantized vibrational states of the acceptor are specifically included, has been expressed^{18,19,22,23} as

$$k = \sqrt{\frac{\pi}{\hbar^2 \lambda_s k_B T}} |T_{\text{ab}}|^2 \sum_{m=0}^{\infty} \frac{e^{-S} S^m}{m!} \exp \left[\frac{-(\lambda_s + \Delta G^0 + m h \nu)^2}{4 \lambda_s k_B T} \right] \quad (4)$$

$$S = \frac{\lambda V}{\hbar \nu} \quad (5)$$

where S is the vibrational-electronic coupling strength, T_{ab} is

the coupling matrix element, and the total reorganization energy is expressed in terms of the solvational (λ_s) and vibrational (λ_v) components. The ΔG° value used in equation 4 was obtained from the assigned relative electron transfer exothermicity for compounds 1–7 ($\Delta G^\circ_{\text{rel}}$, Table 2) and an added shift correction

$$\Delta G^\circ = \Delta G^\circ_{\text{rel}} + \Delta G_{\text{shift}} \quad (6)$$

The ΔG_{shift} term contains coulomb and ion-pairing corrections and any other systematic errors in the assignment of the electron transfer exothermicity. The effect of the ΔG_{shift} correction is to shift the origin of the relative ΔG° scale. The present kinetic analysis assumes no repopulation of the porphyrin excited singlet via the charge-separated state which may be close in energy (for example, compound 1). Work in progress will assess the validity of this assumption.

Using a value of $\Delta G_{\text{shift}} = 0.48$ eV, equation (4) gives a curve of the form shown in Fig. 2 for benzene with the following parameters: $T_{\text{ab}} = 8 \text{ cm}^{-1}$, $\lambda_v = 0.3$ eV, $\lambda_s = 0.3$ eV, and $h\nu = 1,500 \text{ cm}^{-1}$. The choice for $h\nu$ and λ_v are reasonable for the reduction of a benzoquinone and the previously obtained value of T_{ab} . The parameters obtained from the data are in accord with estimates of $h\nu = 1,500 \text{ cm}^{-1}$ and $\lambda_v = 0.3$ eV calculated from an average force constant for the CO stretches of a quinone^{24,25} and a semiquinone (approximated by a phenolic CO stretch), bond length changes^{26,27} between C=O and C—O (~ 0.2 Å), and the extent of spin delocalization onto the semiquinone ring obtained from electron paramagnetic resonance studies²⁸.

In summary, the classical and semi-classical expressions qualitatively describes the rising side of the curve. The fluctuations observed in the insensitive region may be fine structure present in the electron transfer rates for this series that is not adequately described by these theories and may require alternative theoretical interpretations. The ΔG range would have to be extended to more negative values to see whether or not 'inverted behaviour' is obtained. This series examines exothermicity effects on photon-induced electron transfer rates uncomplicated

by major electronic perturbations, changes in outer sphere reorganization energies, or poorly-defined donor-acceptor distances, and should provide critical experimental measurements for further theoretical refinements.

A.H.Z. J.J.H. and P.B.D. thank the NSF for support. We thank Professor R. Marcus for helpful discussions, Professor F. Anson's group for assistance with the electrochemistry, and J. M. Warman for critical reading of the manuscript.

Received 6 October 1986; accepted 9 March 1987.

1. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. *Nature* **318**, 618–624 (1985).
2. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. *J. molec. Biol.* **180**, 385–398 (1984).
3. Leland, B. A. *et al. J. phys. Chem.* **89**, 5571–5573 (1985).
4. Joran, A. D., Leland, B. A., Geller, G. G., Hopfield, J. J. & Dervan, P. B. *J. Am. chem. Soc.* **106**, 6090–6092 (1984).
5. Warman, J. M. *et al. Nature* **320**, 615–616. (1986).
6. Miller, J. R., Calcatera, L. T. & Close, G. L. *J. Am. chem. Soc.* **106**, 3047–3049 (1984).
7. Wasielewski, M. R., Niemczyk, M. P., Svec, W. A. & Pewitt, E. B. *J. Am. chem. Soc.* **107**, 1080–1082 (1985).
8. Irvine, M. P., Harrison, R. J., Beddard, G. S., Leighton, P. & Sanders, J. K. M. *Chem. Phys.* **104**, 315–324 (1986).
9. Rehm, D. & Weller, A. *Israel J. Chem.* **8**, 259–271 (1970).
10. Pasman, P., Mes, G. F., Koper, N. W. & Verhoeven, J. W. *J. Am. chem. Soc.* **107**, 5839–5843 (1985).
11. Schmidt, J. A., Siemiarczuk, A., Weedon, A. C. & Bolton, J. R. *J. Am. chem. Soc.* **107**, 6112–6116 (1985).
12. Harris, D., Johnson, A. W. & Gaete-Holmes, R. *Bioorg. Chem.* **9**, 63–70 (1980).
13. Lambert, W. R., Felker, P. M. & Zewail, A. H. *J. chem. Phys.* **81**, 2217–2232 (1984); Felker, P. M. & Zewail, A. H. *J. chem. Phys.* **82**, 2975–2993 (1985).
14. Marcus, R. A. *J. chem. Phys.* **24**, 966–978 (1956).
15. Hopfield, J. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3640–3644 (1974).
16. Marcus, R. A. & Sutin, N. *Biochem. biophys. Acta* **811**, 265–322 (1985).
17. Gunner, M. R., Robertson, P. E. & Dutton, P. L. *J. phys. Chem.* (in the press).
18. Jortner, J. *J. chem. Phys.* **64**, 4860–4867 (1976).
19. Jortner, J. *J. Am. chem. Soc.* **102**, 6676–6686 (1980).
20. Buhks, E. & Jortner, J. *FEBS Lett.* **109**, 117–120 (1980).
21. Hopfield, J. J. in *Protein Structure: Molecular and Electronic Reactivity* (eds Austin, R. H. *et al.*) (Springer, New York, in the press).
22. Ulstrup, J. & Jortner, J. *J. chem. Phys.* **63**, 4358–4368 (1975).
23. Marcus, R. A. *Discuss. Faraday Soc.* **74**, 7 (1982).
24. Anno, T. & Sadô, A. *Bull. chem. Soc. Japan* **31**, 734–739 (1958).
25. Baudet, M. J., Berthier, G. & Pullman, B. *J. chim. Phys.* **54**, 282–284 (1957).
26. S. M. Swingle, *J. Am. chem. Soc.* **76**, 1409 (1954).
27. Bratoz, M. S. Besnainou, M. S. *J. chim. Phys.* **56**, 555–562 (1959).
28. Higasi, K., Baba, H. & Rembaum, A. *Quantum Organic Chemistry* 242–246 (Interscience, New York, 1965).

Oscillating selection on Darwin's finches

H. Lisle Gibbs* & Peter R. Grant†

* Department of Biology, The University of Michigan, Ann Arbor, Michigan 48109–1048, USA

† Department of Biology, Princeton University, Princeton, New Jersey 08544, USA

An important goal in the study of evolution is to determine the occurrence, causes and possible micro-evolutionary consequences of selection in natural populations^{1–3}. Darwin's finches (*Geospizinae*) are suitable organisms for investigation because their morphological traits are highly heritable⁴, and they live in a climatically variable environment (Galápagos Islands)^{5,6}. It has been suggested that selection fluctuates in direction and intensity, favouring different morphological optima in different years⁷, because strong annual variation in rainfall causes changes in food supply composition^{7,8}. This suggestion has been supported in part by studies of the medium ground finch, *Geospiza fortis*, on the island of Daphne Major, which have shown that large adult size is favoured under drought conditions, when the overall food supply is low and large hard seeds are disproportionately abundant^{8,9}. Here we document a reversal in the direction of selection following the opposite climatic extreme, and demonstrate the connection between oscillating selection and fluctuations in food supply.

From December 1982 until July 1983, exceptionally heavy and prolonged rain fell on Daphne Major^{10,11}, in association with the most severe El Niño event (an oceanic disturbance) of the century¹². A total of 1,359 mm of rain was recorded, ten

times the previously recorded maximum of 137 mm in 1978. Birds bred continuously for eight months^{10,11}, instead of the usual one to three months, and by the time the rains ceased both food level (seeds) and bird density were exceptionally high. We examined the survival of individually ringed and measured birds over the next two years when rainfall was low (53 mm, 1984) to very low (4 mm, 1985). During this time seed biomass steadily declined but was nevertheless higher than in the dry years of 1977 and 1982.

We measured directional selection on adults by comparing the sizes of survivors and non-survivors. Results are given in Table 1 under two headings. Selection gradients (β) represent the direct effects of selection on each trait and are measured as the partial regression coefficients of relative fitness (survival) on each trait¹¹. Selection differentials (s) measure the combined direct effects and indirect effects of selection caused by phenotypic correlations between traits, and are calculated as the difference in the mean value of the trait before and after selection, here measured in standard deviation units¹.

Small adults survived better than large ones in the two years following the exceptionally wet El Niño year (Table 1, Fig. 1); selection differentials for all measured traits were negative and often significant. These are in marked contrast to selection on all traits in the opposite direction under the contrasting conditions of drought and low overall food supply (Fig. 1).

Selection was consistent over sexes and cohorts. For example, all selection differentials for males and females were negative, although not all were significantly different from zero. Males, the larger sex, were apparently subject to stronger selection than females. Selection differentials for overall body size (principal

Table 1 Standardized selection differentials (s) and gradients (β) for adult *G. fortis* from 1984 to the beginning of 1986

Character	1984		Periods of selection 1985		1984-85 combined	
	β	s	β	s	β	s
Weight	-0.12 ± 0.07	-0.12^{**}	-0.03 ± 0.08	-0.08	-0.09 ± 0.10	-0.18^{**}
Wing	-0.07 ± 0.06	-0.09^*	$0.15 \pm 0.07^*$	0.03	0.11 ± 0.08	-0.06
Tarsus	0.03 ± 0.06	-0.06	-0.06 ± 0.06	-0.09	-0.09 ± 0.07	-0.17^{**}
Bill length	0.10 ± 0.06	-0.03	-0.10 ± 0.07	-0.05	0.08 ± 0.09	-0.09
Bill depth	0.13 ± 0.09	-0.08^*	-0.02 ± 0.11	-0.10^*	0.03 ± 0.13	-0.18^{**}
Bill width	$-0.21 \pm 0.09^*$	-0.12^{**}	-0.15 ± 0.10	-0.11^{**}	$-0.29 \pm 0.12^*$	-0.21^{***}
PC I		-0.12^{**}		-0.10^{**}		-0.21^{***}
PC II		0.03		0.00		0.03
N	314 (0.61)	320-391	316 (0.56)	318-341	489 (0.36)	496-537
r^2	0.06*	—	0.05*	—	0.05**	—

Values of β are shown $\pm$ s.e.m. Before each analysis data were $\log_e$ -transformed and standardized to have zero mean and unit variance. Principal component (PC) scores were extracted from a correlation matrix calculated from all individuals for which measurements of all six univariate traits were available. PC I is interpreted as a measure of overall body size because all characters load positively on it, whereas PC II is interpreted as a shape index in terms of bill size relative to tarsus length¹⁵. N is the number of birds alive at the beginning of the period, and varies according to the number of measurements taken. The proportion surviving is given in parentheses. The 1984-85 combined sample includes birds born in 1984. Significance of selection differentials was assessed by t -tests comparing survivors with those who disappeared (* $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$).

component I, PCI in Table 1) were almost twice as large for males (-0.41 , $P < 0.001$) as for females (-0.22 , $P < 0.05$).

Significant β values identify the targets of direct selection among the set of intercorrelated morphological traits⁴ subject to analysis. No single trait, with the possible exception of bill width, was an obvious target of selection (Table 1). In one year selection weakly favoured increased wing length, but no net selection occurred because of a positive correlation with overall size which was selected in the opposite direction. These results differ from those during droughts and conditions of low food supply, when individual targets and directions of selection could be clearly identified (weight and bill depth selected to increase, bill width selected to decrease⁹). Thus we interpret the strong differentials and weak gradients to indicate selection on overall body size, as manifested by highly significant negative selection differentials for PCI scores (Table 1).

Large adult size is favoured when food is scarce because the supply of small and soft seeds is depleted first, and only those birds with large bills can crack open the remaining large and hard seeds^{3,8,9,13}. In contrast, small adult size is favoured in years following very wet conditions, possibly because the food

supply is dominated by small soft seeds. In 1984 and 1985, the proportion of total seed biomass made up by small soft seeds was 21% to 80%¹¹, which is two to ten times greater than the previous maximum (8% in May 1978) following a normal wet year. Despite an initially large standing crop of seeds, finch mortality was high and correlated with size, possibly because large birds had difficulty finding enough large seeds. As has been found in other years^{8,13-15}, finches seen feeding on large seeds were significantly larger in all traits except bill length, and in PCI scores, than those feeding solely on small seeds (t -tests, $P \leq 0.05$) in 1984 and 1985 ($N = 108-124$ birds). Thus morphological constraints on foraging behaviour may underlie the fluctuating selection on body size under the contrasting feeding conditions induced by droughts and El Niño events.

In contrast to adults, juveniles were not subjected to directional selection on body size or bill size during the first six to nine months following their birth in 1983 and 1984. Juveniles were not present in 1985 because no successful breeding occurred in this year. Small size is favoured among juveniles when small seeds are scarce following a breeding season¹⁶.

The chief implication of the results of this long-term study is that natural populations subject to the effects of rare climatic perturbations are not at a fixed equilibrium state demographically, phenotypically, or genetically. Selection in opposite directions on the two sexes¹⁷⁻²⁰, on correlated traits within a sex^{9,19}, on different life-history stages of both sexes^{16,21}, or on the same age and sex group in different years and generations, may result in weak overall stabilizing selection on morphological traits over a period greater than a single generation^{3,4,22}.

We thank Philip de Maynadier, Steven Latta, Robert Podolsky, Dan Rosenberg, Jonathan Weiland and David Wiggins for assistance in the field, and James Fry, Rosemary Grant, Philip Myers, Robert Payne and Trevor Price for advice. This study was conducted with permission and support from the Dirección General de Desarrollo Forestal, Quito, the National Parks Service, Galápagos, and the Charles Darwin Research Station. It was funded by grants from the NSF. H.L.G. was supported by a Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarship for part of the study.

Received 30 January; accepted 24 March 1987.

1. Lande, R. & Arnold, S. J. *Evolution* **37**, 1210-1226 (1983).
2. Endler, J. A. *Natural Selection in the Wild* (Princeton University Press, 1986).
3. Grant, P. R. *Ecology and Evolution of Darwin's Finches* (Princeton University Press, 1986).
4. Boag, P. T. *Evolution* **37**, 877-894 (1983).
5. Grant, P. R. & Boag, P. T. *Auk* **97**, 227-244 (1980).
6. Grant, P. R. *Am. Orn. Un. Monogr.* **36**, 471-483 (1985).

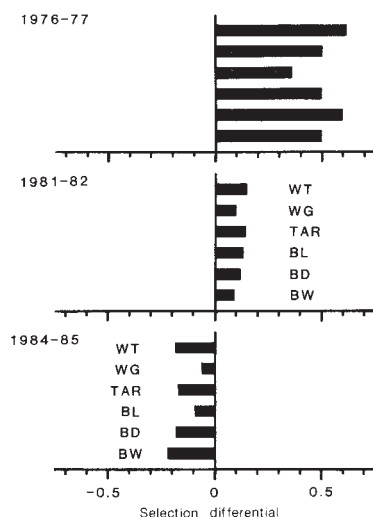


Fig. 1 Selection differentials for traits of adult *G. fortis* during two years of drought and low food supply, 1976-77 and 1981-82 (ref. 9), and during the years that followed the El Niño event of 1982-83. Traits measured are shown as follows. WT, weight; WG, wing length; TAR, tarsus size; BL, bill length; BD, bill depth; BW, bill width.

7. Grant, P. R., Grant, B. R., Smith, J. N. M., Abbott, I. J. & Abbott, L. K. *Proc. natn. Acad. Sci. U.S.A.* **73**, 257-261 (1976).
8. Boag, P. T. & Grant, P. R. *Science* **214**, 82-85 (1981).
9. Price, T. D., Grant, P. R., Gibbs, H. L. & Boag, P. T. *Nature* **309**, 787-789 (1984).
10. Gibbs, H. L., Grant, P. R. & Weiland, J. *Auk* **101**, 872-874 (1984).
11. Gibbs, H. L. & Grant, P. R. *Ecology* **68** (in the press).
12. Cane, M. A. *Science* **222**, 1189-1195 (1983).
13. Price, T. D. *Ecology* **68** (in the press).
14. Grant, P. R. *Anim. Behav.* **29**, 785-793 (1981).
15. Boag, P. T. & Grant, P. R. *Biol. J. Linn. Soc. Lond.* **22**, 243-287 (1984).
16. Price, T. D. & Grant, P. R. *Evolution* **38**, 483-494 (1984).
17. Fleischer, R. C. & Johnston, R. F. *Nature* **298**, 747-749 (1982).
18. Price, T. D. *Am. Nat.* **123**, 500-518 (1984).
19. Grant, B. R. *Evolution* **39**, 523-532 (1985).
20. Monaghan, P. & Metcalfe, P. *Evolution* **40**, 1096-1099 (1986).
21. Schluter, D. & Smith, J. N. M. *Evolution* **40**, 221-231 (1986).
22. Grant, P. R. & Price, T. D. *Am. Zool.* **21**, 795-811 (1981).

NMDA receptors in the visual cortex of young kittens are more effective than those of adult cats

T. Tsumoto, K. Hagihara, H. Sato & Y. Hata

Department of Neurophysiology, Institute of Higher Nervous Activity, Osaka University Medical School, Kitaku, Osaka, 530 Japan

Acidic amino acids, such as glutamate and aspartate, are thought to be excitatory transmitters in the cerebral neocortex and hippocampus¹⁻⁸. Receptors for these amino acids can be classified into at least three types on the basis of their agonists. Quisqualate-preferring receptors and kainate-preferring receptors are implicated in the mediation of synaptic transmission in many regions including the hippocampus^{9,10} and visual cortex¹¹, whereas N-methyl-D-aspartate (NMDA)-preferring receptors are thought to be involved in modulating synaptic efficacy, for example in long-term potentiation, a form of synaptic plasticity in the hippocampus¹²⁻¹⁴. In the visual cortex of the cat and monkey, it is well established that synaptic plasticity, estimated by susceptibility of binocular responsiveness of cortical neurons to monocular visual deprivation, disappears after the 'critical' period of postnatal development¹⁵⁻¹⁷. Here we report that during the critical period in young kittens, a selective NMDA-receptor antagonist blocks visual responses of cortical neurons much more effectively than it does in the adult cat. This suggests that NMDA receptors may be involved in establishing synaptic plasticity in the kitten visual cortex.

Nine adult cats and seven kittens, aged between 4 and 8 weeks and thus at a stage when the synaptic plasticity of the visual cortex is highest^{15,16}, were used. Experimental procedures for preparing and maintaining the animals, visual and electrical

stimulations and extracellular single unit recordings were as detailed previously⁵. During recordings the animals were anaesthetized with a gas mixture of 70% N₂O and 30% O₂, and paralysed with gallamine triethiodide (Flaxedil). The heart rate was continuously monitored. If it was suddenly changed by noxious stimuli, 0.1-0.5% halothane was added to the gas to maintain an adequate depth of anaesthesia. Three- or four-barrelled micropipettes for ionophoretic application of drugs contained 2-amino-5-phosphonovaleic acid (APV, 50 mM, pH 8.0), a selective NMDA antagonist^{9,10}, and kynurenic acid (50 mM, pH 8.0), a broad-spectrum excitatory amino-acid antagonist^{5,18,19}. In addition, the pipettes contained L-glutamate (0.2 M, pH 8.0), NMDA (50 mM, pH 8.0), quisqualate (20 mM, pH 8.0) and kainate (20 mM, pH 8.0).

In adult cats, 69 cells recorded from the primary visual cortex were tested with drugs. As in a previous report⁵, ionophoretic application of kynurenic acid effectively suppressed visual responses in 71% of the cells tested (Fig. 1a). The application of APV, on the other hand, significantly suppressed visual responses in only 33% of the cells (Fig. 1b and Table 1). The current intensity for APV was set either at, or 10-20 nA above, the intensity at which APV blocked completely excitations induced by the application of NMDA. Also, APV was ineffective in suppressing responses elicited by single electrical shocks applied to the dorsal lateral geniculate nucleus (LGN) in 77% of the cells tested whereas kynurenic acid blocked the responses, particularly those with latencies shorter than 2 ms, in 84% of the cells (and see ref. 20).

In the kittens, however, APV suppressed visual responses in 71% of the cortical cells tested (Fig. 2a and Table 1). Usually, APV suppressed visual responses with approximately the same intensity of current as kynurenic acid, but the suppression by APV was less complete than that by kynurenic acid. With this range of currents, APV completely blocked excitations induced by NMDA but not by quisqualate and kainate, as in the adult. In double-spike responses evoked by LGN stimulation (Fig. 2b), the initial responses were found relatively resistant to APV but the secondary spikes were highly susceptible.

A summary of the results is shown in Table 1. In the adult, the proportion of cells suppressed by APV was only 33% whereas it was 71% in the kittens. The difference in these values is statistically significant (χ^2 -test, $P < 0.001$). The laminar distribution of the cells in the kittens was not markedly different from that in the adult. In the 44 suppressed cells of the kittens, the mean intensity of current ($\pm$ s.d.) required for APV to exert its full action was 43 ± 17 nA, which is significantly (t -test, $P < 0.005$) smaller than that of the adult (65 ± 22 nA, number of cells, $n = 23$). Also, the minimum intensity of currents for NMDA needed to excite cells in the kittens (33 ± 20 nA, $n = 18$)

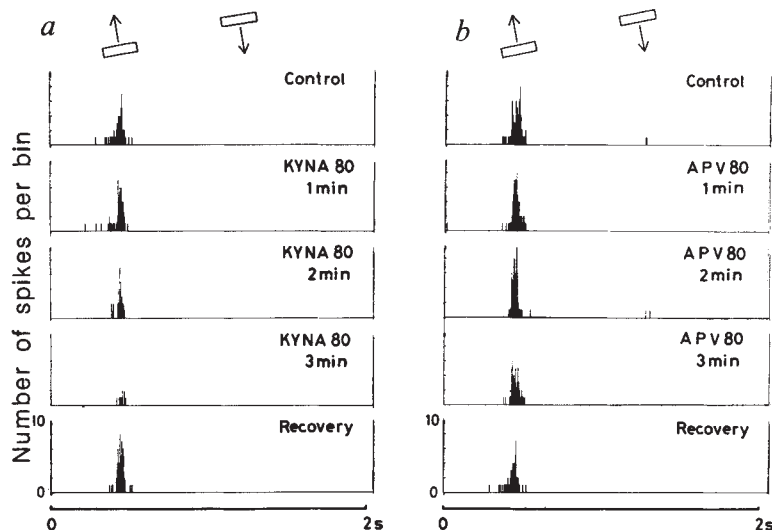


Fig. 1 Effects of an ionophoretic application of kynurenic acid and 2-amino-5-phosphonovaleic acid (APV) on visual responses of a cortical neuron recorded from an adult cat. *a*, Suppressive effects of kynurenic acid (KYNA); *b*, lack of effect of APV. Shown are peristimulus time histograms of spikes recorded extracellularly to a light slit moving back and forth over the receptive field. Direction of motion and orientation of the slit are indicated at the top. Twenty sweeps for each histogram; bin width 4 ms. Stimulus speed was 17.2 deg s^{-1} . Top row, control histograms cumulated without application of any drug; second to fourth rows, histograms cumulated during application of the drug indicated. Numbers at the right of drug name, current intensity used (nA); numbers below drug name, time after starting the ejection of each drug. Bottom row, control histograms cumulated after the end of the drug application.

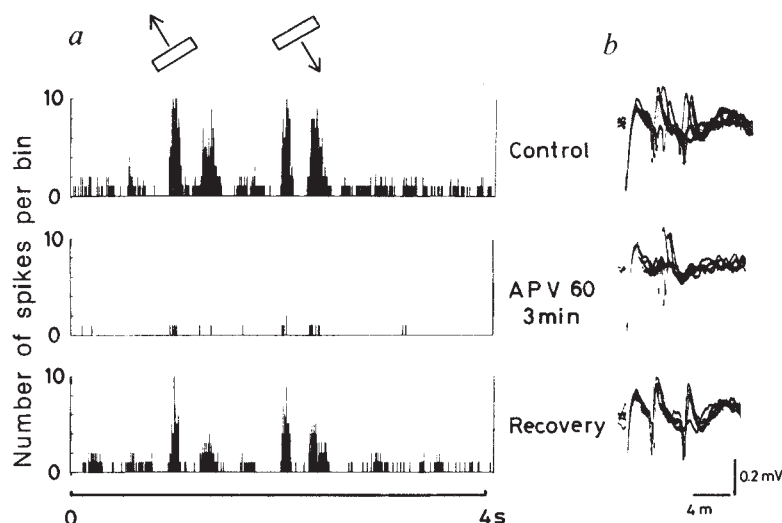


Fig. 2 Suppression of visual responses (*a*) and electrically-elicited responses (*b*) of a kitten cortical neuron by APV application. *a*, Each histogram shows peristimulus time histogram of spikes to an optimally oriented slit moving back and forth over the receptive field. Direction of motion and orientation of the slit are indicated at the top. Ten sweeps for each histogram, bin width 8 ms. Stimulus speed was 8.6 deg s^{-1} . Top, control histogram cumulated without application of any drug. Middle, histogram cumulated during application of APV with 60 nA. Time after starting the ejection of APV is also indicated. Bottom, control histogram cumulated after stopping the APV application. *b*, Responses to single shock stimulation of LGN before (top), during (middle) and after (bottom) the application of APV. Superimposition of five sweeps.

was significantly ($P < 0.01$) smaller than that in the adult ($57 \pm 28 \text{ nA}$, $n = 48$, including the data from ref. 5). However, the minimum current intensity for quisqualate ($53 \pm 19 \text{ nA}$, $n = 17$) and for kainate ($47 \pm 21 \text{ nA}$, $n = 16$) to excite cells in the kittens were not significantly different from those in the adult (54 ± 28 , $n = 30$ and $52 \pm 25 \text{ nA}$, $n = 23$, respectively, including the data from ref. 5).

The present results strongly suggest that NMDA receptors participate in synaptic responses of visual cortical neurons in young kittens but not appreciably in adults. There is another possibility, however, namely that immature cortical neurons in the developing cortex are very susceptible to drugs applied ionophoretically, so that the suppression by APV in kittens could be a general depressant effect, irrespective of its selective antagonistic action. This possibility seems unlikely, however, because first, APV did not suppress quisqualate- and kainate-induced excitations of cortical cells with the current intensity with which visual responses and NMDA-induced excitations were clearly blocked, and second, cortical neurons in the kittens had about the same sensitivity to quisqualate and kainate as those in the adult. From the latter results it also seems unlikely that the difference in sensitivity of cortical neurons to APV between the kittens and adults might be due to a difference in spatial diffusion of the ionophoretically applied drugs or in their access to receptors. These considerations do not necessarily imply, however, that non-NMDA receptors, such as quisqualate and kainate receptors, are not involved in visual responses in kittens. In fact, the suppression of visual responses by APV was not so complete as that by kynurenic acid in most of the cells. Furthermore, the initial spikes of LGN-induced responses, which could be blocked by kynurenic acid, were relatively resistant to APV whereas the secondary spikes were easily blocked. In the rat cortex, hippocampus and thalamus, NMDA receptors become active only when the cell membrane is depolarized²¹⁻²⁴. We suggest that NMDA receptors participate in gen-

erating visual responses, which consist of high-frequency discharge of spikes, following initial depolarization by non-NMDA receptors in kitten cortical neurons, and that this action of NMDA receptors becomes subthreshold or disappears as the non-NMDA receptors mature.

In the rat hippocampus, NMDA receptors are believed to be crucial in synaptic plasticity, probably controlling entry of Ca^{2+} into neurons which initiates Ca^{2+} -dependent biochemical processes that underlie long-lasting changes in synaptic function^{25,26}. In the visual cortex of young kittens, NMDA receptors operate very effectively, so that weak excitatory inputs to cortical neurons through immature non-NMDA receptors may be potentiated, as discussed above. Therefore, it seems reasonable to assume that operation of NMDA receptors may also induce the Ca^{2+} entry into immature cortical neurons and consequently trigger the processes underlying synaptic plasticity in the kitten visual cortex. Thus, NMDA receptors may not only enhance visual responses of immature cortical neurons but also may induce activity-dependent synaptic modifications in the developing visual cortex.

This work was supported by a Grant-in-Aid for Special Project Research of Plasticity of Neural Circuits from the Japanese Ministry of Education, Science and Culture.

Received 12 January; accepted 14 April 1987.

Table 1 Number of cells with visual responses which were significantly suppressed by APV

	Suppressed significantly	Not suppressed	Total
Adults	23 (33%)	46 (67%)	69
Kittens	44 (71%)	18 (29%)	62

Suppression was judged significant if the total number of spikes in the optimal response was $<50\%$ of the average number of spikes in the control responses before and after the APV application. Results from cells which were lost before completing control histograms after stopping the drug application were discarded.

- Hayashi, T. *Keio J. Med.* **3**, 183-192 (1954).
- Krnjevic, K. & Phillis, J. W. *J. Physiol., Lond.* **165**, 274-304 (1963).
- Clark, R. M. & Collins, C. G. S. *J. Physiol., Lond.* **262**, 383-400 (1976).
- Baughman, R. W. & Gilbert, C. D. *J. Neurosci.* **1**, 427-439 (1981).
- Tsumoto, T., Masui, H. & Sato, H. *J. Neurophysiol.* **55**, 469-483 (1986).
- Nadler, J. V., Vaca, K., White, W. F., Lynch, G. S. & Cotman, C. W. *Nature* **260**, 538-540 (1976).
- Nitsh, C. in *Amino Acid Neurotransmitters* (eds Defeudis, F. V. & Mandel, P.) 97-104 (Raven, New York, 1981).
- Storm-Mathisen, J. in *Glutamate as a Neurotransmitter* (eds Di Chiara, G. & Gessa, G. L.) 43-55 (Raven, New York, 1981).
- Watkins, J. C. & Evans, R. H. *Rev. Pharmac. Tox.* **21**, 165-204 (1981).
- McLennan, H. *Prog. Neurobiol.* **20**, 251-271 (1983).
- Tsumoto, T., Masui, H. & Sato, H. in *Excitatory Amino Acid Transmission* (eds Hicks, T. P., Lodge, D. & McLennan, H.) 389-396 (Liss, New York, 1987).
- Bliss, T. V. P. & Lomo, T. *J. Physiol., Lond.* **232**, 331-356 (1973).
- Collingridge, G. L., Kehl, S. J. & McLennan, H. *J. Physiol., Lond.* **334**, 19-31 (1983).
- Harris, E. W., Ganong, A. H. & Cotman, C. W. *Brain Res.* **323**, 132-137 (1984).
- Hubel, D. H. & Wiesel, T. N. *J. Physiol., Lond.* **206**, 419-436 (1970).
- Blakemore, C. & Van Sluyters, R. C. *J. Physiol., Lond.* **237**, 195-216 (1974).
- Von Noorden, G. K. *Invest. Ophthalmol. Vis. Sci.* **12**, 721-726 (1973).
- Perkins, M. N. & Stone, T. W. *Brain Res.* **247**, 184-187 (1982).
- Ganong, A. H., Lanthorn, T. H. & Cotman, C. W. *Brain Res.* **273**, 170-174 (1983).
- Hagihara, K., Tsumoto, T., Sato, H. & Hata, T. in *Excitatory Amino Acid Transmission* (eds Hicks, T. P., Lodge, D. & McLennan, H.) 409-412 (Liss, New York, 1987).
- Dingledine, R. *J. Physiol., Lond.* **343**, 385-405 (1983).
- Thompson, A. M., West, D. C. & Lodge, D. *Nature* **313**, 479-481 (1985).
- Salt, T. E. *Nature* **322**, 263-265 (1986).
- Herron, C. E., Lester, R. A. J., Coan, F. J. & Collingridge, G. L. *Nature* **322**, 265-268 (1986).
- Fagg, G. E., Foster, A. C. & Ganong, A. H. *Trends Pharmac. Sci.* **7**, 357-363 (1986).
- MacDermott, A. B., Mayer, M. L., Westbrook, G. L., Smith, S. J. & Barker, J. L. *Nature* **321**, 519-522 (1986).

Functional acetylcholine receptor in PC12 cells reacts with a monoclonal antibody to brain nicotinic receptors

P. J. Whiting*, R. Schoepfer*, L. W. Swanson*†, D. M. Simmons*† & J. M. Lindstrom*

* The Salk Institute for Biological Studies, San Diego, California 92138, USA

† Howard Hughes Medical Institute, San Diego, California 92138, USA

The use of a variety of probes for neuronal nicotinic acetylcholine receptors¹⁻⁶ has indicated that there are sites on neurons that bind the acetylcholine receptor antagonist α -bungarotoxin¹, but do not regulate cation channels in response to the binding of acetylcholine^{6,7}. Sites with high binding affinity for nicotine and for α -bungarotoxin also show different distributions in brain¹. A monoclonal antibody (mAb35) raised against acetylcholine receptors from *Electrophorus* electric organ has been used to purify receptors from chick brain⁸. These receptors do not bind α -bungarotoxin but have a high affinity for nicotine⁹ and antibodies raised against them block the function of acetylcholine receptors in ciliary ganglia¹⁰. We have raised a monoclonal antibody (mAb270) against acetylcholine receptors from chicken brain (P.J.W., R. Liu, S. Shimasaki, F. Esch, B. Morley & J.M.L., manuscript in preparation), which has been used to purify a similar receptor from rat brain¹¹. Here we show that this antibody identifies a functional nicotinic acetylcholine receptor in rat-neuron-like PC12 cells which is probably not encoded by the cDNA clone (λ PCA48) previously suggested as a candidate for the acetylcholine receptor gene⁵. Additionally, we show that mAb270 binds to the same areas of rat brain as nicotine.

The structure and properties of nicotinic acetylcholine receptors (nAChRs) from brains of chickens and rats are summarized in Table 1. The nAChRs from brains of both chickens and rats appear to be composed of only α - and β -subunits rather than the α -, β -, γ - and δ -subunits which compose nAChRs from electric organ and muscle. The α -subunits of nAChRs from brain share antigenic determinants with α -subunits of nAChRs from electric organ and muscle^{8,11} but it is the β -subunits of brain AChRs that contain the acetylcholine binding site¹², rather than the α -subunits, as is the case for nAChRs from electric organ and muscle¹³. Two nAChR subtypes have been identified in chicken brain (P.J.W. *et al.*, in preparation). Both have the

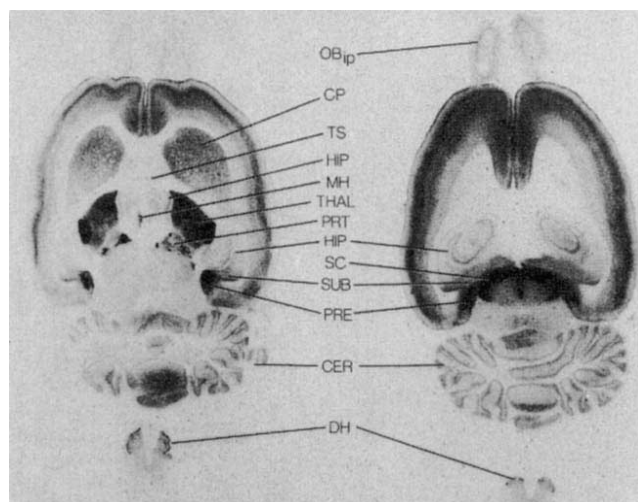


Fig. 1 Photomicrographs showing the distribution of ¹²⁵I-labelled mAb270-binding sites in two horizontal sections through the rat brain and spinal cord. Abbreviations: CER, cerebellar granular layer; CP, caudoputamen; DH, spinal cord dorsal horn; HIP, hippocampus and dentate gyrus; MH, medial habenula; OBip, olfactory bulb inner plexiform layer; PRE, presubiculum; PRT, pretectal region; SC, superior colliculus; SUB, subiculum; THAL, thalamus; TS, triangular nucleus of the septum.

Methods. For immunohistochemistry, the animals were decapitated and the brain was removed and frozen with liquid nitrogen. Cryostat sections (20 μ m) were thaw-mounted onto slides and desiccated at 0–4 °C under vacuum overnight. The sections were incubated overnight at 4 °C with 4 nM ¹²⁵I-labelled mAb270 (specific activity $2-4 \times 10^{18}$ c.p.m. mol⁻¹) in 100 mM NaCl, 10 mM sodium phosphate buffer (pH 7.5), 10 mM NaN₃, 10% normal rat serum, and 5% dried milk (Carnation). The slides were then transferred to Coplin jars and rinsed five times over 30 min. with 100 mM NaCl, 10 mM sodium phosphate buffer (pH 7.5) and 10 mM NaN₃ at room temperature. They were further rinsed in three changes of buffer over 3 h on a rocking platform at 4 °C, and were then dried at 37 °C, mounted on cardboard, overlain with an 8 $\times$ 10-inch sheet of Kodak XAR5 film in a cassette, and autoradiographed at room temperature for 12–36 h. Virtually no labelling was observed when the sections were coincubated in 400 nM unlabelled mAb270. Adjacent Nissl-stained sections were used to identify labelled structures.

same high affinity for binding L-nicotine, do not bind α -bungarotoxin (α -BT), and are separable by their binding of monoclonal antibody (mAb) 35 which binds to the subtype with β -subunits of relative molecular mass 59,000 (M_r 59K) and by mAb285 (raised against nAChRs purified from chicken brain) which binds to the subtype with β' -subunits of M_r 75K. To date, only nAChRs with β' -type subunits have been identified in rat brains¹¹. Using mAbs, homologous nAChRs have been identified in human brain¹⁴. It is thought that the subunits of nAChRs from electric organ and muscle¹⁵ and from neurons^{11,12} are encoded by part of an extended gene family which also contains genes for the neuronal α -BT binding component¹⁶.

Each of these proteins may have evolved from a primordial nAChR composed of a single type of subunit which, by a process of repeated gene duplication and mutation along each lineage of this family produced homologous subunits composing nAChRs from muscles, ganglia and brain, and the α -BT binding component of neurons.

Figure 1 shows the binding of ¹²⁵I-labelled mAb270 to sections of rat brain. Detailed studies³¹ show that the pattern of binding closely follows that reported for ³H-nicotine, but is quite different from that for ¹²⁵I-labelled α -BT (ref. 1), as would be expected from the observation that mAb270 can immunoprecipitate >90% of the high-affinity ³H-nicotine binding sites in a

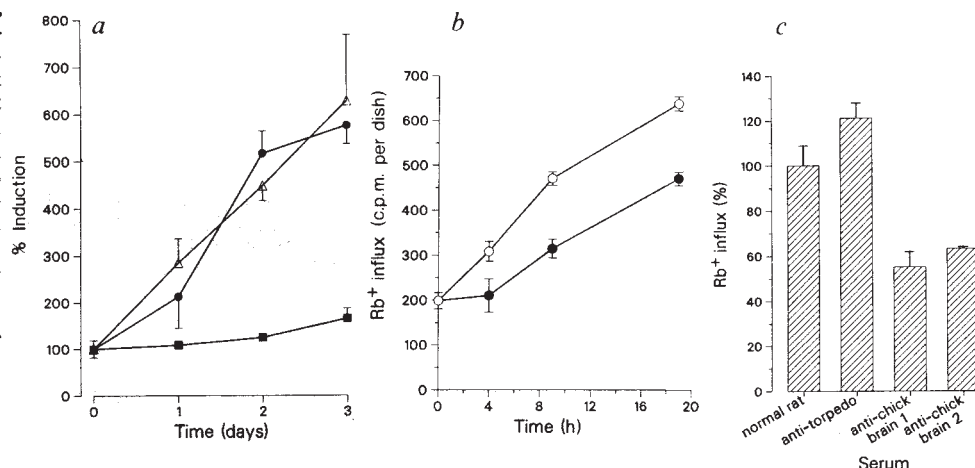
Table 1 Properties of nAChRs from brains of chickens and rats

	Chicken brain*		Rat brain†
Fraction of high-affinity ³ H-nicotine binding sites bound by mAb270	>90%		>90%
nAChR subtypes	$\alpha\beta$	$\alpha\beta'$	$\alpha\beta'$
α -subunit			
M_r	49,000	49,000	51,000
mAb270 binding	+	+	+
β -subunit			
M_r	59,000	75,000	79,000
ACh binding site	+	+	+
K_1 L-nicotine	1.1×10^{-9} M	1.6×10^{-9} M	1.3×10^{-9} M
K_1 α -BT	$>10^{-6}$ M	$>10^{-6}$ M	$>10^{-6}$ M
K_1 atropine	$>10^{-3}$ M	$>10^{-3}$ M	$>10^{-3}$ M

* Data from refs 8, 9 and 12 and P.J.W., R. Liu, S. Shimasaki, F. Esch, B. Morley, and J.M.L., manuscript in preparation.

† Data from refs 9, 11 and 12.

Fig. 2 Binding of mAb270 to a functional nAChR of PC12 cells. **a**, Induction by β NGF of carbamylcholine-stimulated $^{86}\text{Rb}^+$ influx and mAb270 binding sites, but not of α -BT binding sites, in PC12 cells. PC12 cells were plated out on 35-mm dishes and on day 0, medium containing β NGF (50 ng ml^{-1}) was added. The carbamylcholine-stimulated $^{86}\text{Rb}^+$ influx ($\bullet$), ^{125}I -labelled mAb270 binding sites (Δ), and ^{125}I -labelled α -BT binding sites ($\square$) were measured on days 0, 1, 2 and 3. Data are shown as the mean $\pm$ s.d. of triplicate cultures. **b**, Modulation of carbamylcholine-stimulated $^{86}\text{Rb}^+$ influx by mAb270. Cell monolayers were cultured for 20 h in 1 ml medium containing 50 ng ml^{-1} β NGF. Then mAb270 or control mAb164 (which has the same IgG subclass as mAb270, IgG2a, but binds only to nAChRs from *Torpedo*²³) were added to a final concentration of 100 nM, and after 5 h, at time 0, 1 ml of β NGF medium containing a 1/80 dilution of goat anti-rat IgG (prepared by 18% sodium sulphate fractionation of immune goat serum) was added. At time 0, 4, 9 and 19 h the carbamylcholine-stimulated $^{86}\text{Rb}^+$ uptake in cells treated with mAb270 ($\bullet$) and mAb164 ($\circ$) was determined. Results are expressed as mean $\pm$ s.d. of triplicate cultures. **c**, Modulation of carbamylcholine-stimulated $^{86}\text{Rb}^+$ influx by antisera to nAChRs from chicken brain. Cell monolayers were cultured for 4 d in medium containing 50 ng ml^{-1} β NGF and then for 16 h in medium containing a 1/25 dilution of normal rat serum, rat serum raised against nAChRs from *Torpedo* (titre $22\text{ }\mu\text{M}$) and two rat sera raised against immunoaffinity-purified nAChRs from chicken brain (titres $32\text{ }\mu\text{M}$ and $96\text{ }\mu\text{M}$, ref. 8). The carbamylcholine-stimulated $^{86}\text{Rb}^+$ influx was then determined as described above. Results are mean $\pm$ s.d. of triplicate cultures and are expressed considering the $^{86}\text{Rb}^+$ influx in the presence of normal rat serum ($85\text{ nmol per } 30\text{ s per } 8 \times 10^5\text{ cells}$) as 100%.



Methods. PC12J cells²⁴ were maintained in culture in 175-cm² plastic flasks in Iscove's Medium supplemented with a 10% fetal calf serum and 10% horse serum. For assays, cells were replated on polylysine-coated dishes. Uptake of $^{86}\text{Rb}^+$ was determined by the method of Catterall²⁵. Briefly, cell monolayers were washed with 3 ml of assay buffer, with 1 ml of assay buffer containing ouabain, and then incubated for 30 s in ouabain-containing buffer with $2\text{ }\mu\text{Ci ml}^{-1}$ $^{86}\text{Rb}^+$ (NEN) and 5 mM carbamylcholine. Cell monolayers were washed three times with 3 ml assay buffer, scraped off in 2 ml of 0.5 M NaOH, and radioactivity determined on a scintillation counter at 55% efficiency. Nonspecific uptake of $^{86}\text{Rb}^+$ was determined in the absence of carbamylcholine, and subsequently subtracted. To assay ^{125}I -labelled mAb270 binding, mAb270 prepared from rats immunized with immunoaffinity purified chicken brain AChR⁹ was purified by chromatography on DEAE Tris Acryl (Pharmacia) and radioiodinated by a modified chloramine-T method²⁶. Labelling was carried out in culture medium containing 5% horse serum. Cell cultures were washed with 3 ml of medium and then incubated for 30 min at 37 °C in 0.25 ml of medium with or without 1.0 μM non-radioactive mAb270. Then 0.25 ml of 20 nM ^{125}I -labelled mAb270 in culture medium was added and after a further 45 min at 37 °C the cell monolayers were washed three times with 3 ml of culture medium, scraped off in 2 ml of 0.5 M NaOH, and radioactivity was determined by γ -counting. Specific binding of ^{125}I -labelled mAb270 was determined by subtraction of the binding in the presence of 1 μM non-radioactive mAb270. For α -BT binding assays, α -BT was radioiodinated to an initial specific activity of $3\text{--}4 \times 10^{17}\text{ c.p.m. mol}^{-1}$ (ref. 26). Cultures were washed with 3 ml of medium, and then incubated for 30 min at 37 °C in 0.25 ml medium with or without 5 mM carbamylcholine. Then ^{125}I -labelled α -BT (0.25 ml of 20 nM) was added and, after a further 45-min incubation, cell monolayers were washed three times and radioactivity determined by γ -counting. Specific binding of ^{125}I -labelled α -BT was determined by subtracting the binding in 5 mM carbamylcholine.

Triton X-100 detergent extract of rat brains, but does not precipitate the ^{125}I -labelled- α -BT binding sites^{9,11}. The use of ^{125}I -labelled mAb270 allowed autoradiography by exposure for 12 h rather than for six weeks, as use of ^3H -nicotine requires. Labelling in the retina, optic tract and most terminal fields of the optic nerve, such as the superior colliculus, was eliminated by removing an eye, suggesting that nAChRs synthesized in retinal ganglion cells may function presynaptically in areas such as the superior colliculus.

The rat pheochromocytoma cell line PC12 has nAChRs that do not bind α -BT and α -BT binding components which are not nAChRs⁶. The nAChR of PC12 cells is related antigenically to brain nAChRs and reacts similarly with ACh-binding-site affinity labels¹⁷. The PC12 nAChR, in contrast to nAChRs purified from brain⁹, does not have high affinity for ^3H -nicotine. By both gel-filtration assay⁹ and indirect immobilization of detergent-solubilized PC12 nAChRs upon mAb270-Sepharose we could not detect any ^3H -nicotine binding sites, even though sufficient nAChRs (measured in terms of ^{125}I -labelled-mAb270 binding sites on intact cells) were solubilized to have given a readily detectable signal in these assays. Others have reported that PC12 nAChRs do not exhibit high-affinity binding for ^3H -acetylcholine¹⁸.

The function of nAChRs in PC12 cells can be studied conveniently (in contrast to nAChRs in brain) by measuring car-

bamylcholine-induced $^{22}\text{Na}^+$ or $^{86}\text{Rb}^+$ influx^{6,19}. Binding of mAb270 to nAChRs from brain does not inhibit nicotine binding⁹ or directly block carbamylcholine-stimulated $^{86}\text{Rb}^+$ influx into PC12 cells (data not shown). To test whether mAb270 binds to functional nAChRs, we took advantage of the observation that treatment of PC12 cells with β -nerve growth factor (β NGF) causes an increase in nAChRs but not in α -BT binding sites¹⁹. Untreated PC12 cells have $\sim 10,000$ α -BT binding sites per cell, but only 1,000 mAb270 binding sites. After PC12 cells were treated with β NGF, ^{125}I -labelled-mAb270 binding increased in parallel with the number of functional nAChRs (Fig. 2a). On day 3 of culture in β NGF, the $^{86}\text{Rb}^+$ influx was 93 nmol per 8×10^5 cells per 30 s, and there were 7,100 ^{125}I -labelled-mAb270 binding sites per cell and 16,400 ^{125}I -labelled- α -BT binding sites per cell. Neither the number of functional nAChRs nor mAb270 binding changed when PC12 cells were treated with fibroblast growth factor (1–2 ng per ml growth medium), even though this causes morphological differentiation similar to that induced by β NGF, further showing that functional nAChRs and mAb270 binding are precisely co-regulated, as expected if mAb270 binds to the nAChR molecule (data not shown).

To test whether mAb270 binds directly to the functional nAChRs in PC12 cells, we investigated the ability of mAb270 to down-regulate the amount of nAChR. Treatment of muscle cells with antibody against nAChRs cross-links the nAChRs

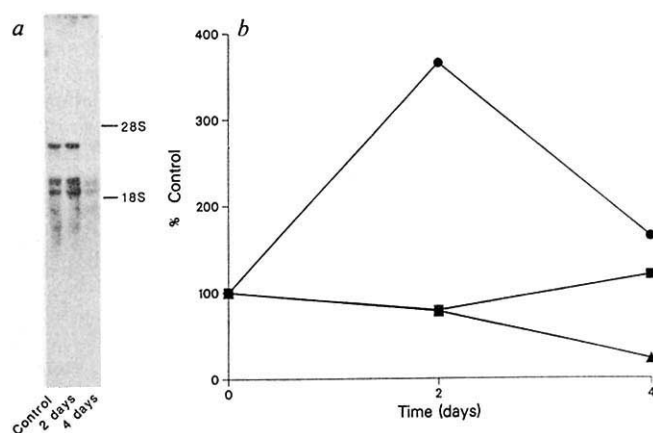


Fig. 3 RNA levels of β NGF-induced PC12 cells. *a*, Northern blot of PC12 RNA probed with insert of λ PCA48. *b*, Slot blot analysis of PC12 RNA probed with inserts of the cDNA clone λ PCA48 (▲), γ -actin (■), and SCG10 (●).

Methods. PC12 cells were cultured in 175-cm² flasks as in Fig. 2. When semiconfluent (day 0) β NGF-containing medium (50 ng ml⁻¹) was added. Cells were collected after 2 or 4 d in culture. Cells from a control flask cultured without β NGF was also collected on day 4. Total cellular RNA was isolated by the guanidine thiocyanate/caesium chloride method²⁷. For the Northern blot, 10 μ g RNA of each preparation was size-fractionated through a 1% agarose gel containing 2.2 M formaldehyde and transferred onto a nitrocellulose filter (BA 85, Schleicher and Schuell) and for the slot blot analysis a serial dilution of each RNA preparation was filtered onto nitrocellulose²⁸. Hybridization was performed in 5 $\times$ SSPE, 5 $\times$ Denhardt's solution, 50% formamide, 0.1% SDS, 150 μ g ml⁻¹ salmon sperm DNA at 42°C with probes for λ PCA48 (ref. 5), γ -actin²⁹ or SCG10 (ref. 29) labelled to high specific activity by primer extension with Klenow polymerase in the presence of [α -³²P]dCTP³⁰. Stringent washing (twice with 1 $\times$ SSPE containing 0.1% SDS, and twice with 0.3 $\times$ SSPE, containing 0.1% SDS at 65°C, 10 min each wash) was followed by autoradiography. The 28S and 18S ribosomal RNA bands were visualized on the nitrocellulose filter by staining with methylene blue. The Northern blot analysis of PC12 RNA probed with λ PCA48 revealed a pattern of RNA species essentially identical to that observed by Boulter *et al.*⁵. Slot blot autoradiograms were quantified on a scanning densitometer (Hoefer Scientific Instruments). Peak area values of the hybridization signal of RNA from cells grown in the absence of β NGF were considered as 100%, and other values expressed relative to this. The RNA species which hybridizes with λ PCA48 decreases in cells grown in the presence of β NGF. Control slot blots of the same RNA preparations probed with a γ -actin cDNA sequence demonstrated no significant change in hybridization signal²⁹. As a further control, when slot blots were probed with the SCG10 cDNA probe (which hybridizes to an RNA species in PC12 cells induced by β NGF²⁹), there was a significant increase in the hybridization signal to RNA from β NGF-induced cells (although less than previously reported using a different subline of PC12²⁹). This confirms differentiation of the PC12 cells at the mRNA level by β NGF.

nAChRs from mouse muscle. This cDNA could encode a subunit of the nAChR, or a subunit of the α -BT binding protein, or some other related protein. The amino-terminal amino-acid sequences which we determined for the two subunits of nAChRs from rat brain were not the same as the deduced N-terminal sequence of λ PCA48 (P.J.W., F. Esch, S. Shimasaki and J.M.L., unpublished data), suggesting that λ PCA48 does not encode a subunit of the brain nAChR bound by mAb270. However we know that nAChRs in PC12 cells differ subtly from brain nAChRs in the selectivity of their reaction with affinity labelling reagents^{12,17}, and their much lower affinity for acetylcholine¹⁸. Also, λ PCA48 does hybridize *in situ* with rat brain sections in some of the same areas which bind mAb270 (ref. 21). Such data are difficult to interpret, however, because *in situ* hybridization localizes messenger RNA in cell bodies, whereas binding of mAb270 and nicotine localizes protein which may be most concentrated at the ends of axons such as those in the superior colliculus (Fig. 1). Also, because *in situ* hybridization was conducted at reduced stringency, hybridization with λ PCA48²¹ might detect mRNA for an nAChR, an α -BT binding component, or both.

Figure 3 shows that mRNA which hybridizes with λ PCA48 is not induced in PC12 cells by treatment with β NGF. This phenomenon could be explained if the appearance of nAChRs induced by β NGF (Fig. 2*a*) is regulated at the level of translation, protein assembly, or transport. However, in muscle a similar increase in the amount of nAChR which occurs after denervation is due primarily to increased transcription²². Because β NGF treatment substantially increases nAChR and mAb270 binding but does not significantly influence the number of α -BT binding sites (Fig. 2) or the quantity of the RNA species which hybridizes with λ PCA48 (Fig. 3), it is possible that λ PCA48 encodes a subunit of the α -BT binding component, or another as yet unidentified member of this gene family.

In conclusion, mAb270 clearly recognizes a functional nAChR in PC12 cells and its use has permitted the localization, purification and characterization of brain nAChRs^{9,11,12}. In future studies it may prove useful in identifying the cDNAs coding for the subunits of these nAChRs, and reveal other interesting features of their structure and function.

We thank Dr David Schubert for the PC12J cell line, for β NGF and fibroblast growth factor, and for helpful comments. We also thank Dr James Boulter and his colleagues for the λ PCA48 cDNA clone and Dr David Anderson for gifts of the γ -actin and SCG10 cDNA clones. We gratefully acknowledge the technical assistance of Douglas Schneider and John Cooper. This work was supported by grants to J.M.L. from the NIH, the Muscular Dystrophy Association, the Alexander S. Onassis Public Benefit Foundation, the US Army and the Los Angeles and California Chapters of the Myasthenia Gravis Foundation. P.W. was supported by the Carl M. Pearson Postdoctoral Fellowship of the Muscular Dystrophy Association.

Received 11 February; accepted 8 April 1987.

- Clarke, P. B. S., Schwartz, R. D., Paul, S. M., Pert, C. B. & Pert, A. *J. Neurosci.* **5**, 1307-1315 (1985).
- Chiappinelli, V. A. *Brain Res.* **277**, 9-21 (1983).
- Halvorsen, S. W. & Berg, D. K. *J. Neurosci.* **6**, 3405-3412 (1986).
- Swanson, L. W. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 4532-4536 (1983).
- Boulter, J. *et al. Nature* **319**, 368-374 (1986).
- Patrick, J. & Stallcup, W. B. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4689-4692 (1977).
- Egan, T. M. & North, R. A. *Neuroscience* **19**, 565-571 (1986).
- Whiting, P. J. & Lindstrom, J. M. *Biochemistry* **25**, 2082-2093 (1986).
- Whiting, P. J. & Lindstrom, J. M. *J. Neurosci.* **6**, 3061-3069 (1986).
- Stollberg, J., Whiting, P. J., Lindstrom, J. M. & Berg, D. K. *Brain Res.* **378**, 179-182 (1986).
- Whiting, P. J. & Lindstrom, J. *Proc. natn. Acad. Sci. U.S.A.* **84**, 595-599 (1987).
- Whiting, P. J. & Lindstrom, J. *FEBS Lett.* **213**, 55-60 (1987).
- Popot, J. L. & Changeux, J. P. *Physiol. Rev.* **64**, 1162-1239 (1984).
- Whiting, P. J., Cooper, J. & Lindstrom, J. *J. Neuroimmun.* (in the press).
- Noda *et al. Nature* **305**, 818-823 (1983).
- Conti-Tronconi, B., *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 5208-5212 (1985).
- Leprince, P. *Biochemistry* **22**, 5551-5556 (1983).
- Kemp, G. & Morley, B. J. *FEBS Lett.* **205**, 265-268 (1986).
- Amy, C. M. & Bennett, E. L. *J. Neurosci.* **3**, 1547-1553 (1983).
- Heinemann, S., Bevan, S., Kullberg, R., Lindstrom, J. & Rice, J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3090-3094 (1977).

and speeds their endocytosis and lysosomal destruction by a process termed 'antigenic modulation'²⁰. Similarly, mAb270 (Fig. 2*b*) and antisera to brain nAChRs (Fig. 2*c*) cause antigenic modulation of nAChRs in PC12 cells. For modulation by mAb270, it was necessary also to add anti-immunoglobulin G (IgG) to enhance cross-linking. This was not necessary with the polyclonal sera, probably owing to multivalent binding of several antibody species to each nAChR allowing more efficient cross-linking of nAChRs.

Recently Boulter *et al.*⁵ reported that screening a complementary DNA library from PC12 cells (not treated with β NGF) with a cDNA for the α -subunit of nAChRs from mouse muscle detected a cDNA clone (λ PCA48) which exhibited 47% deduced amino-acid sequence identity with the α -subunit of

21. Goldman, D., Simmons, D., Swanson, L. W., Patrick, J. & Heinemann, S. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4076-4080 (1986).
22. Merlie, J., Isenberg, K., Russell, S. & Sanes, J. *J. Cell Biol.* **99**, 332-355 (1984).
23. Gullick, W. J. & Lindstrom, J. M. *Biochemistry* **22**, 3312-3320 (1983).
24. Hatanaka, H. *Dev. Brain Res.* **6**, 243-250 (1983).
25. Catterall, W. A. *J. biol. Chem.* **250**, 1776-1781 (1975).
26. Lindstrom, J., Einarsen, B. & Tzartos, S. *Meth. Enzym.* **76**, 452-460 (1981).
27. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294-5299 (1979).
28. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
29. Anderson, D. J. & Axel, R. *Cell* **42**, 649-662 (1985).
30. Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **137**, 266-267 (1984).
31. Swanson, L., Simmons, D., Whiting, P. & Lindstrom, J. *J. Neurosci.* (in the press).

A second human interleukin-2 binding protein that may be a component of high-affinity interleukin-2 receptors

Mitchell Dukovich*, Yuji Wano*, Le thi Bich Thuy†, Paul Katz‡, Bryan R. Cullen§, John H. Kehrl|| & Warner C. Greene*

The Metabolism Branch* and Laboratory of Biochemistry†,

National Cancer Institute, Laboratory of Immunoregulation||, National Institute of Allergy and Infectious Diseases,

National Institutes of Health, Bethesda, Maryland, USA

§ Department of Molecular Genetics, Hoffmann-LaRoche Inc., Nutley, New Jersey, USA

‡ Division of Rheumatology, Immunology and Allergy, Georgetown University School of Medicine, Washington, DC, USA

Although activated human T and B lymphocytes express both high-affinity and low-affinity membrane receptors for interleukin-2 (IL-2)¹⁻³, the structural features that distinguish these receptors have remained unresolved. The high-affinity receptors appear to mediate IL-2 induced T cell growth^{1,2} and internalization of IL-2^{4,5}, whereas no function has yet been ascribed to the low-affinity receptors. The Tac antigen is an IL-2 binding protein⁶⁻¹¹ of relative molecular mass 55,000 (M_r 55K) that participates in the formation of both high- and low-affinity receptors². But Tac complementary DNA transfection¹²⁻¹⁷ and membrane fusion¹⁸ studies have suggested that additional T-cell components are required to produce high-affinity IL-2 receptors. In this study, we report the identification of a second human IL-2 binding protein that (1) has an M_r of ~70K, (2) lacks reactivity with the anti-Tac antibody, (3) binds IL-2 with intermediate affinity and (4) is present on the surface of resting T cells, large granular lymphocytes (natural killer cells), and certain T and B cell lines in the absence of the Tac antigen. Chemical crosslinking of ¹²⁵I-labelled IL-2 bound to high-affinity IL-2 receptors produces labelling of both the p70 protein and the Tac antigen and the anti-Tac antibody blocks the crosslink detection of both of these proteins. Expression of Tac cDNA in a T cell line expressing the p70 protein, but lacking both Tac and high-affinity receptors, results in the reconstitution of high-affinity IL-2 receptors in these cells. Together, these findings suggest that the high-affinity human IL-2 receptor may be a membrane complex composed of at least the p70 protein and Tac antigen.

Purified recombinant ¹²⁵I-labelled IL-2 was bound and cross-linked with disuccinimidylyl suberate (DSS) to varying cell lines under conditions favouring an interaction with the high-affinity receptors (20 pM final concentration of ligand). Using either HUT 102B2 cells or PHA-activated normal T cells, two major ¹²⁵I-labelled IL-2 crosslinked protein bands were detected under high-affinity conditions (Fig. 1a, b). The smallest protein (A) corresponded in size (M_r 66-69K) to that predicted for the Tac antigen (M_r 50-55K) associated with IL-2 (M_r 15.5K). The second major band (B) migrated with an apparent M_r of 83-89K

suggesting either the crosslinking of IL-2 to a M_r 68-74K protein or the binding of a second IL-2 molecule to the Tac antigen. Two larger and less prominent bands (C and D) were detected in some but not all experiments (Fig. 1a). IL-2-dependent murine CTLL cells, which display both high- and low-affinity mouse IL-2 receptors, exhibited a similar pattern of ¹²⁵I-labelled IL-2 crosslinking; however, band B appeared larger (M_r 95K) and bands C and D were prominent (Fig. 1a). In contrast, under high-affinity conditions, no crosslinked proteins were detected on Tac cDNA-transfected mouse fibroblasts (L-Trans 3) that display only low-affinity IL-2 receptors¹³ or Jurkat T cells that lack high-affinity IL-2 binding sites (Fig. 1a).

Addition of a 200-fold molar excess of unlabelled IL-2 or anti-Tac antibody blocked ¹²⁵I-labelled IL-2 crosslink detection of both species (Fig. 1c). In contrast, neither the 7G7/B6 antibody, which reacts with the Tac antigen but does not interfere with IL-2 binding¹⁹, nor the control IgG_{2a} UPC10 antibody had any effect. These results are similar to the pattern of high-affinity IL-2 crosslinking to HUT 102B2 or activated normal T cells recently described by Kuo *et al.*²⁰ and Sharon *et al.*²¹ or with forskolin-induced cells from the 'natural killer-like' YT cell line described by Robb *et al.*²².

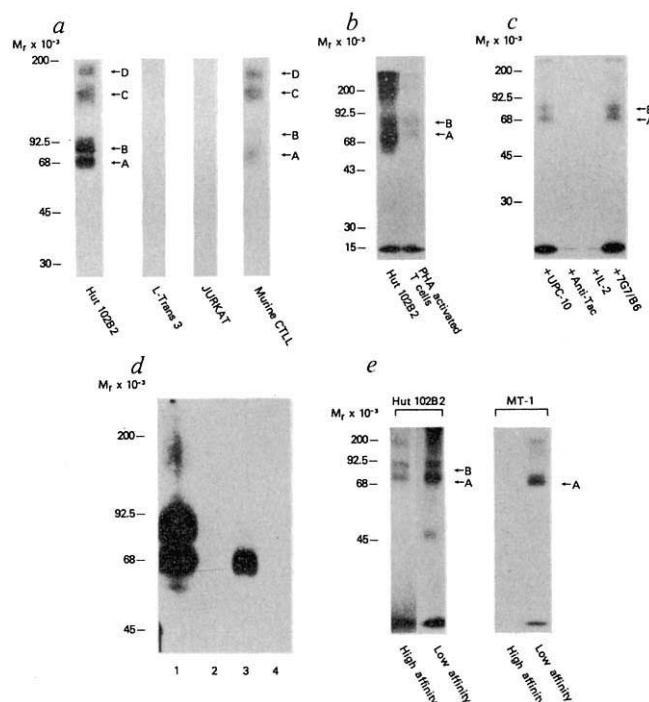
To investigate the identity of bands A and B, the crosslinked proteins were immunoprecipitated with a rabbit heteroantiserum specific for the Tac antigen¹². Although it quantitatively immunoprecipitated band A, the Tac-specific heteroantiserum failed to immunoprecipitate any of band B despite over-exposure of the gel (Fig. 1d). Similar results were obtained with the 7G7/B6 monoclonal antibody (data not shown). Preimmune control rabbit serum or purified anti-BSA antibodies did not immunoprecipitate either of the crosslinked proteins. These findings suggested that band A corresponded to the Tac antigen crosslinked to ¹²⁵I-labelled IL-2 whereas band B was produced by the crosslinking of single molecule of ¹²⁵I-labelled IL-2 to a 68-74K protein which we shall refer to as p70. In further support of a difference in these proteins, Sharon *et al.*²¹ and Robb *et al.*²² have demonstrated that proteolytic cleavage of each of these crosslinked complexes produces a different series of peptide fragments.

To compare the pattern of crosslinking when IL-2 was associated with the high- or low-affinity forms of the receptor, ¹²⁵I-labelled IL-2 was added at a final concentration of 50 pM (high affinity) or 5 nM (low affinity) to either HUT 102B2 or MT-1 cells (Fig. 1d). Although both bands A and B were present in HUT 102B2 cells, the intensity of band A relative to band B was greater under low-affinity conditions. This result was consistent with an interaction of the ¹²⁵I-labelled IL-2 with the numerous low-affinity Tac binding sites (~180,000 sites per cell) present on these cells². In contrast, crosslinking to HTLV-I-infected MT-1 cells which express low affinity, but not high affinity, IL-2 receptors⁵ revealed the presence of band A but not band B (Fig. 1e). These findings indicate that MT-1 cells lack the p70 protein and is correlated with the absence of high-affinity IL-2 receptors.

These crosslinking and immunoprecipitation results do not distinguish between the possibilities that the p70 protein can bind IL-2 directly or, alternatively, is crosslinked to IL-2 by virtue of close physical proximity to the Tac protein. To investigate potential IL-2 binding by the p70 protein, various lymphoid populations reported to bind or respond to IL-2 in the absence of the Tac antigen were studied. Gibbon ape leukaemia virus-transformed MLA-144 T cells have been described as containing IL-2 binding sites¹ but lacking reactivity with the anti-Tac antibody. This antibody, however, immunoprecipitates the 55K gibbon equivalent of the Tac protein that is present on the surface of activated gibbon T lymphoblasts (W.C.G., unpublished data). In addition, M. Tsudo & T. A. Waldmann have found evidence of a 75K IL-2-binding protein on the surface of these cells (personal communication). No IL-2 receptors were identified on the MLA-144 T cells under high-affinity ¹²⁵I-

* Present address: Howard Hughes Medical Institute, Duke University School of Medicine, PO Box 3705, Durham, North Carolina 27710, USA.

Fig. 1 Crosslinking of ^{125}I -labelled IL-2 under high- and low-affinity binding conditions and immunoprecipitation analysis of crosslinked proteins. For high-affinity crosslinking, ^{125}I -labelled IL-2 ($37.1 \mu\text{Ci } \mu\text{g}^{-1}$, NEN) was added at a final concentration of 20–50 pM and crosslinked with disuccinimidyl suberate (DSS, $100 \mu\text{g ml}^{-1}$) as previously described²⁰. Cell membranes were prepared, solubilized in 1% SDS, and analysed on SDS-polyacrylamide gels. **a**, High-affinity crosslinking of ^{125}I -labelled IL-2 to 25×10^6 HUT 102B2, L-Trans-3, Jurkat or murine IL-2-dependent cytotoxic T cells (CTL). Crosslinked proteins are designated in order of increasing size as bands A, B, C and D. **b**, High-affinity crosslinking to 50×10^6 HUT 102B2 or normal peripheral blood T cells activated with E-PHA ($10 \mu\text{g ml}^{-1}$) for 48 h. Radioactivity at the dye front corresponds to bound but uncrosslinked ^{125}I -labelled IL-2 dissociated from the receptor after boiling in 1% SDS. **c**, Specificity of high-affinity ^{125}I -labelled IL-2 crosslinking to HUT 102B2 cells. Aliquots of 10×10^6 cells were incubated with ^{125}I -labelled IL-2 (20 pM final concentration) in the presence of a 200-fold molar excess of UPC10 monoclonal antibody (IgG_{2a}- κ , Litton), anti-Tac, purified recombinant unlabelled IL-2 or the 7G7/B6 monoclonal antibody¹⁹ followed by DSS crosslinking and SDS-PAGE analysis. **d**, Immunoprecipitation of ^{125}I -labelled IL-2 crosslinked proteins from HUT 102B2 cells. Membranes from 100×10^6 HUT 102B2 crosslinked to ^{125}I -labelled IL-2 under high-affinity binding conditions (50 pM final concentration) were solubilized and immunoprecipitated as previously described⁷. Lane 1, crosslinked proteins not immunoprecipitated; lane 2, preimmune normal rabbit serum; lane 3, Tac-specific rabbit heteroantiserum; lane 4, anti-bovine serum albumin specific heteroantibody. Autoradiograph shown represents 14-day exposure. **e**, Comparison of high- and low-affinity ^{125}I -labelled IL-2 crosslinking to HUT 102B2 and MT-1 cells. Aliquots of 10×10^6 HUT 102B2 or MT-1 cells were incubated with ^{125}I -labelled IL-2 at a final concentration of 50 pM (high affinity) or 5 nM (low affinity) followed by DSS crosslinking and SDS-PAGE analysis.



labelled IL-2 crosslinking conditions (Fig. 2a). In contrast, under low-affinity conditions, a crosslinked protein comigrating with band B in the absence of band A was detected (Fig. 2a). IL-2 binding to the p70 protein on MLA-144 T cells was blocked by unlabelled IL-2 but not by the anti-Tac antibody (Fig. 2b). In this and some other experiments, band B resolved as an apparent triplet of proteins (M_r 83K, 85K and 92K). At present it is unclear whether this finding reflects variable post-translational modification of p70, limited degradation during processing, or IL-2 crosslinking to different proteins. A similar crosslinked complex of M_r 83K and 90K was also demonstrated on YT cells by Robb *et al.*²²; digestion with glycosidases reduced this complex to a doublet of M_r 72K and 76K indicating that it contains carbohydrate²².

The SKW6.4 B cell line responds to high concentrations of IL-2 with increased immunoglobulin secretion but lacks detectable Tac antigen expression²³. High-affinity crosslinking of ^{125}I -labelled IL-2 to these cells revealed no bands, but low-affinity crosslinking demonstrated the presence of small quantities of the p70 protein in the absence of the Tac protein (Fig. 2c).

The lack of Tac antigen expression in the MLA-144 T cells and SKW6.4 B cells was further confirmed in RNA blotting studies. In contrast to HUT 102B2 and MT-1 cells, the MLA-144 and SKW6.4 B cells did not contain detectable Tac mRNA (Fig. 2d). These results argue against the possibility that the p70 protein corresponds to an altered form of the Tac protein lacking reactivity with anti-Tac.

The number and apparent affinity of the p70 IL-2 binding sites on the MLA-144 T cells and the SKW6.4 B cells was determined. Scatchard plots of ^{125}I -labelled IL-2 binding illustrate that MLA-144 T cells displayed a single class of IL-2 binding sites (4,300 sites per cell) having an apparent K_d of 640 pM (Fig. 2e). SKW6.4 B cells contained 920 binding sites per cell and an apparent K_d of 850 pM (Fig. 2f). Anti-Tac did not inhibit IL-2 binding to either of these cell lines (data not shown). These results suggest that the p70 protein corresponds

to a second IL-2 receptor which binds IL-2 with an intermediate affinity compared with the previously described high- and low-affinity forms of the IL-2 receptor^{1–3}.

Normal peripheral blood lymphoid populations were also examined for expression of the p70 protein. Freshly isolated, unstimulated peripheral blood T cells have been reported to proliferate in response to high doses of IL-2 in the absence of other activation signals^{24,25}. High affinity ^{125}I -labelled IL-2 crosslinking to purified T cells (>99% positive for the CD3 antigen) lacking the Tac antigen (<0.1% reactive with anti-Tac or 7G7/B6) revealed no crosslinked bands (Fig. 3a). Low-affinity crosslinking to these cells, however, demonstrated the presence of the p70 protein. Crosslinking was blocked by unlabelled IL-2 but not by the anti-Tac antibody (Fig. 3a). ^{125}I -labelled IL-2 binding assays confirmed the presence of intermediate-affinity IL-2 binding sites in this cell population.

Large granular lymphocytes (LGLs) were similarly studied for the presence of the p70 protein as natural killer (NK) cell function can be augmented in these cells by the addition of large quantities of IL-2 even in the presence of the anti-Tac antibody²⁶. As shown in Fig. 3b, purified LGLs (>90% reactive with the Leu 11a antibody²⁷) expressed the p70 protein detected by low-affinity crosslinking of ^{125}I -labelled IL-2 but lacked the Tac antigen.

The apparent affinity of the p70 IL-2 binding sites on the MLA 144 T cells was lower than that characteristic of the high-affinity IL-2 receptors present on normal activated T or HUT 102B2 cells. The high-affinity-crosslink detection of both the Tac antigen and p70 protein coupled with capacity of anti-Tac to block the labelling of each protein suggested that high-affinity IL-2 receptor expression might require the combined presence of these proteins, perhaps assembled in a receptor complex. To test this hypothesis, the Tac cDNA was transfected into the MLA-144 T cells and resultant IL-2 receptor affinity was measured. Protoplasts containing plasmids with the Tac cDNA promoted by the early region of the cytomegalovirus (pBC140IL-

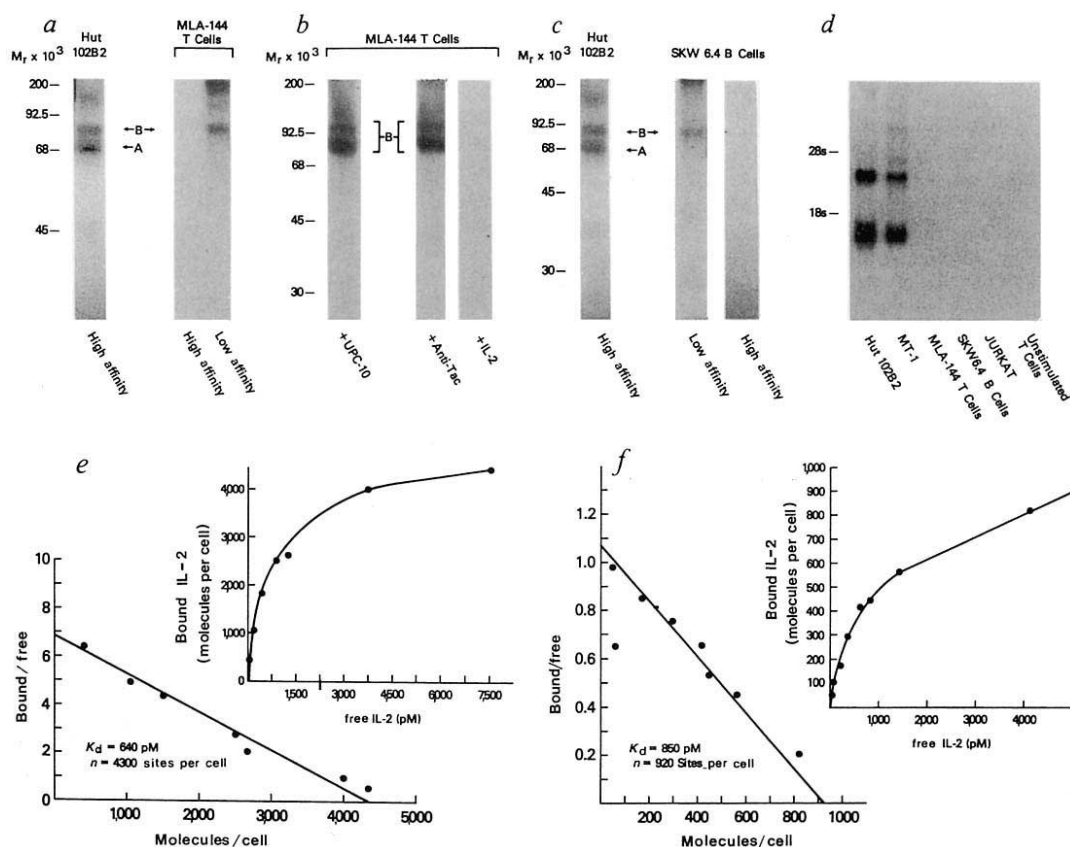
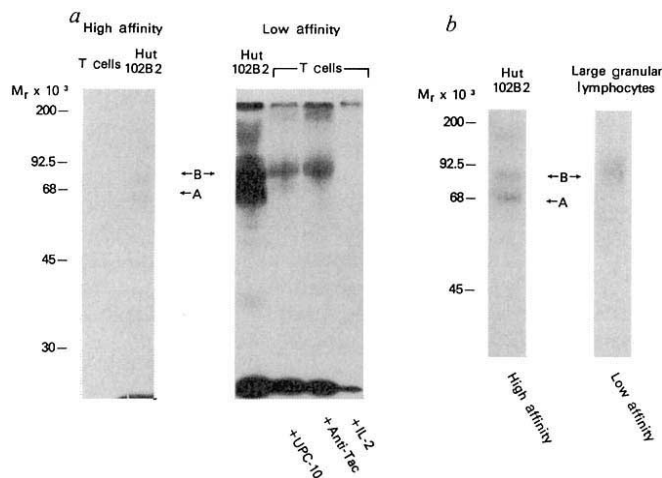


Fig. 2 *a*, High- (50 pM) and low-affinity (5 nM) 125 I-labelled IL-2 crosslinking to MLA-144 T cells (10×10^6 cells) compared with high-affinity crosslinking to HUT 102B2 (10×10^6 cells). *b*, Low-affinity (5 nM) crosslinking of 125 I-labelled IL-2 to MLA-144 T cells (20×10^6) in the presence of a 200-fold molar excess of UPC-10 antibody, anti-Tac antibody, or unlabelled purified recombinant IL-2. *c*, High- (50 pM) and low- (5 nM) affinity 125 I-labelled IL-2 crosslinking to SKW6.4 B cells (30×10^6) compared with high-affinity crosslinking to HUT 102B2 cells (10×10^6). *d*, Northern blot analysis of RNA from HUT 102B2, MT-1, MLA-144, SKW6.4, Jurkat and unstimulated normal peripheral blood T cells. $10 \mu\text{g}$ of RNA from each of these cell cultures were size fractionated on a formaldehyde-agarose gel, transferred to nitrocellulose, hybridized with 32 P-labelled Tac cDNA, and washed as previously described¹². *e*, 125 I-labelled IL-2 binding assay on MLA-144 T cells. Aliquots of 2×10^6 cells were incubated with various quantities of 125 I-labelled IL-2 (2.5–7,500 pM final concentration) for 2 h at 4 °C in the presence or absence of a 500-fold molar excess of unlabelled IL-2 as previously described^{1,2}. Specifically bound and free 125 I-labelled IL-2 was measured and the data analysed by Scatchard plots. The derived values for the dissociation constant (K_d) and number of receptor sites per cell are given. *f*, 125 I-labelled IL-2 binding assay on SKW6.4 B cells. Assay conditions as for *e*.

Fig. 3 *a*, 125 I-labelled IL-2 crosslinking under high- and low-affinity binding conditions to unstimulated normal peripheral blood T cells. Peripheral blood mononuclear cells were obtained by leukopheresis and Tac-negative T cells were purified as previously described³¹ at 4 °C in serum-free conditions to preclude inadvertent cellular activation. Immunofluorescent staining revealed that >99% of the cells expressed the CD3 surface marker while <0.1% of the cells reacted with anti-Tac. Aliquots of 150×10^6 T cells or 10×10^6 HUT 102B2 cells were crosslinked to 125 I-labelled IL-2 at a final concentration of 50 pM (high affinity) or 5 nM (low affinity) in the presence or absence of a 200-fold molar excess of UPC-10, anti-Tac, or unlabelled purified recombinant IL-2. *b*, Low-affinity 125 I-labelled IL-2 crosslinking to large granular lymphocytes (LGLs). LGLs (40×10^6) were purified from leukopheresis samples of peripheral blood mononuclear cells in three steps including (1) Ficoll/Hypaque-density gradient centrifugation, (2) passage over nylon wool columns, and (3) depletion of the nylon wool nonadherent cells of CD3⁺, CD5⁺, and Ia⁺ cells by ox red cell rosette formation. More than 90% of the cells in the resultant population reacted with the Leu 11a antibody which recognizes the IgG Fc receptor present on natural killer cells²⁷. These cells were incubated with 125 I-labelled IL-2 (5 nM final concentration, low affinity), crosslinked, and analysed as previously described. HUT 102B2 cells (10×10^6) were bound and crosslinked to 125 I-labelled IL-2 at a final concentration of 50 pM (high affinity).



2R)¹⁷ were prepared and fused to MLA-144 cells using polyethylene glycol²⁸. Indirect immunofluorescent staining with anti-Tac revealed Tac antigen expression on 10–25% of the cell population transfected with the Tac cDNA but not on cells transfected with protoplasts containing the pBC140 plasmid lacking a cDNA insert.¹²⁵I-labelled IL-2 binding assays were performed on each of these cell populations (Fig. 4a). MLA-144 cells transfected with the control plasmid displayed only intermediate-affinity receptors (K_d 1.3 nM, 4,200 sites per cell). But MLA-144 cells transfected with the Tac cDNA expression plasmid (pBC140-IL2R) displayed both relatively high-affinity (K_d 50 pM, 1,000 sites per cell) and intermediate-affinity (K_d 1.2 nM, 6,000 sites per cell) IL-2 binding sites. In three additional experiments not shown, expression of Tac cDNA produced high-affinity sites with K_d s of 30–75 pM. These reconstituted high-affinity binding sites involved the Tac antigen as

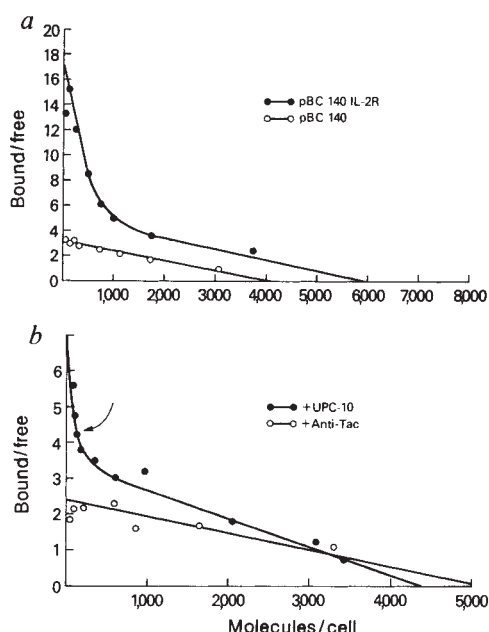


Fig. 4 Expression of the Tac cDNA in MLA-144 T cells. *a*, Scatchard plots of ¹²⁵I-labelled IL-2 binding to: ●, MLA-144 T cells transfected with the Tac cDNA (pBC140 IL-2R); upper part of curve, K_d = 50 pM, n = 1,000 sites per cell; lower part of curve, K_d = 1.2 nM, n = 6,000 sites per cell; ○, MLA-144 cells transfected with the control pBC140 plasmid K_d = 1.3 nM, n = 4,200 sites per cell. No correction has been made for the percentage of Tac⁺ cells in the transfected cultures in the calculation of receptor number. This experiment was performed three times with similar results. *b*, MLA-144 T cells were transfected with the pBC140 IL-2R expression plasmid as described and ¹²⁵I-labelled IL-2 binding was measured in the presence of 1,000-fold molar excess of ●, the UPC-10 antibody; upper part of curve, K_d = 70 pM, n = 500 sites per cell; lower part of curve, K_d = 1.25 nM, n = 4,400 sites per cell; ○, anti-Tac monoclonal antibody, K_d = 2.0 nM, n = 5,100 sites per cell. This experiment was performed twice with similar results.

Methods. For transfection of MLA-144 T cells, protoplasts were prepared by the method of Yoakum *et al.*²⁸ from HB101 *Escherichia coli* transformed with the expression plasmid (pBC140IL-2R) containing the complete protein coding region of the Tac cDNA (880-bp *NciI* fragment of pIL-2R3; ref. 12) cloned in the sense orientation 3' to the immediate early region promoter of the human cytomegalovirus²⁷. Control protoplasts containing the pBC140 plasmid lacking a cDNA insert were similarly prepared. MLA-144 T cells were fused with the protoplasts at a ratio of 1:10⁴ in the presence of 48% polyethylene glycol (Merck) dissolved in RPMI-1640. After fusion, the cells were cultured for 2–4 days in RPMI 1640 containing 20% fetal calf serum. Viable cells were purified by density gradient centrifugation on Ficoll-Hypaque, and ¹²⁵I-labelled IL-2 radioreceptor binding assays were performed as previously described^{1,2}.

the addition of anti-Tac blocked the high-affinity but not the intermediate-affinity component of binding (Fig. 4b). In contrast, the control UPC-10 antibody had no effect on high-affinity binding. In studies of the YT cell line, induction of Tac antigen expression by forskolin stimulation similarly shifts most of the intermediate-affinity p70 binding sites to a high-affinity status²².

These studies demonstrate the existence of a novel 70K IL-2 receptor that is distinct from the Tac protein. Our crosslinking and transfection studies suggest that the high-affinity form of the IL-2 receptor may correspond to a membrane complex composed of the Tac and p70 proteins. But a direct demonstration of a ternary complex of the p70, Tac and IL-2 proteins has not yet been achieved perhaps reflecting the relative inefficiency of the crosslinking reactions and the requirement for two internal covalent bonds to stabilize such a complex. Recently, Robb *et al.* have assembled evidence for different, but perhaps overlapping, binding sites on the IL-2 molecule for the p70 and Tac proteins²². Certainly, a membrane complex composed of two proteins which interact with different IL-2 epitopes provides a plausible mechanism for the generation of high-affinity binding.

The p70 protein has been detected on at least some resting T cells, large granular lymphocytes and certain T and B cell lines in the absence of the Tac antigen. Thus, expression of the p70 protein on these cells provides a possible explanation of their previously considered 'paradoxical' responsiveness to high concentrations of IL-2 and argues that the p70 protein alone may be capable of signal transduction. This possibility is further supported by the recent observation that the p70 protein, unlike the Tac antigen, can mediate rapid internalization of membrane-bound IL-2 (ref. 29). Unlike many receptors composed of multiple subunits, the p70 and Tac proteins appear able to be transported to and displayed on the cell surface in the absence of the other protein. Additional study of the biochemical and molecular properties of the p70 protein may provide important insights into the structure and function of the high-affinity IL-2 receptor.

We thank Ms Gail Whalen for assistance in the preparation of the large granular lymphocyte populations, Dr I. Miyoshi, Kochi Medical School for the MT-1 cells, Drs T. A. Waldmann and D. L. Nelson, NIH for the anti-Tac and 7G7B6 monoclonal antibodies respectively and the Cetus Corporation for purified recombinant IL-2.

Note added in proof: Since submission of this paper, M. Tsudo *et al.*³² and K. Teshigawara *et al.*³³ have reported similar IL-2 binding properties of an M_r 75 K protein and proposed its participation in the formation of high-affinity IL-2 receptors.

Received 17 November 1986; accepted 2 April 1987.

- Robb, R. J., Munk, A. & Smith, K. A. *J. exp. Med.* **154**, 1455–1474 (1981).
- Robb, R. J., Greene, W. C. & Rusk, C. M. *J. exp. Med.* **160**, 1126–1146 (1984).
- Lowenthal, J. W., Zubler, R. H., Nabholz, M. & MacDonald, H. R. *Nature* **315**, 669–672 (1985).
- Weismann, A. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 1463–1466 (1986).
- Fujii, M., Sugamura, K., Sano, K., Nakai, M., Sugita, K. & Hinuma, Y. *J. exp. Med.* **163**, 550–562 (1986).
- Uchiyama, T., Broder, S. A. & Waldmann, T. A. *J. Immun.* **126**, 1393–1397 (1981).
- Leonard, W. J. *et al. Nature* **300**, 267–269 (1982).
- Leonard, W. J., Depper, J. M., Robb, R. J., Waldmann, T. A. & Greene, W. C. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6957–6961 (1983).
- Robb, R. J. & Greene, W. C. *J. exp. Med.* **158**, 1332–1337 (1983).
- Wano, Y., Uchiyama, T., Fukui, K., Maeda, M., Uchino, H. & Yodoi, J. *J. Immun.* **132**, 3005–3010 (1984).
- Leonard, W. J. *et al. J. biol. Chem.* **260**, 1872–1880 (1985).
- Leonard, W. J. *et al. Nature* **311**, 626–631 (1984).
- Greene, W. C., Robb, R. J., Svetlik, P. B., Rusk, C. M., Depper, J. M. & Leonard, W. J. *J. exp. Med.* **162**, 363–368 (1985).
- Sabe, H. *et al. Molec. biol. Med.* **2**, 379–396 (1984).
- Hatakeyama, M. *et al. Nature* **318**, 467–470 (1985).
- Kondo, S. *et al. Nature* **320**, 75–77 (1986).
- Wano, Y., Cullen, B., Pfeffer, N., Svetlik, P. A. & Greene, W. C. *Molec. biol. Med.* (in the press).
- Robb, R. J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3992–3996 (1986).
- Rubin, L. A., Kurman, C. C., Biddison, W. E., Goldman, N. D. & Nelson, D. L. *Hybridoma* **4**, 91–102 (1985).
- Kuo, L.-M., Rusk, C. M. & Robb, R. J. *J. Immun.* **137**, 1544–1551 (1986).
- Sharon, M., Klausner, R. D., Cullen, B. R., Chizzonite, R. & Leonard, W. J. *Science* **234**, 859–863 (1986).
- Robb, R. J., Rusk, C. M., Yodoi, J. & Greene, W. C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2002–2006 (1987).

23. Ralph, P. G. *et al.* *J. Immun.* **133**, 2442-2445 (1984).
24. Le thi Bich Thuy, Lane, C. & Fauci, A. S. *Fed. Proc.* **43**, 1438 (1984).
25. Lifson, J. *et al.* *J. biol. Response Modifiers* **5**, 61-72 (1986).
26. Ortaldo, J. R. *et al.* *J. Immun.* **133**, 779-783 (1984).
27. Lanier, L. L., Le A. M., Phillips, J. H., Warner, N. L. & Babcock, G. F. *J. Immun.* **131**, 1789-1796 (1983).
28. Yoakum, G. H. *et al.* *Science* **222**, 385-389 (1983).
29. Robb, R. J. & Greene, W. C. *J. exp. Med.* **165**, 1201-1206 (1987).
30. Brenner, M. B., Trowbridge, I. S. & Strominger, J. L. *Cell* **40**, 183-190 (1985).
31. Le thi Bich Thuy, Lane, H. C. & Fauci, A. S. *Cell Immun.* **98**, 396-410 (1986).
32. Tsudo, M., Kozak, R. W., Goldman, C. K. & Waldmann, T. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9694-9698 (1986).
33. Teshigawara, K., Wang, H.-M., Kato, K. & Smith, K. A. *J. exp. Med.* **165**, 223-238 (1987).

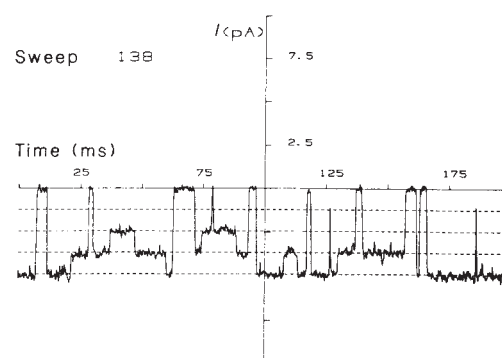


Fig. 1 Record of single-channel activity in a cell-attached patch at a steady pipette holding potential of +140 mV. The pipette solution contained (in mM) 100 KCl, 1.0 MgCl₂, 1.0 CaCl₂ and 10 HEPES, titrated to pH 7.4 with KOH. The bath solution contained (in mM) 97.5 NaCl, 2.5 KCl, 1.0 CaCl₂, 1.0 MgCl₂, and 10 HEPES, titrated to pH 7.4 with NaOH. Single-channel currents were recorded to a video cassette recorder (VCR) following pulse-code modulation and were later analysed with the aid of a mini-computer (low-pass Bessel filter, 1 kHz; sampling frequency 5 kHz). Zero current axis indicates closed channel current, openings downwards (inward current). The dashed lines are drawn at levels of one quarter of fully-open channel current.

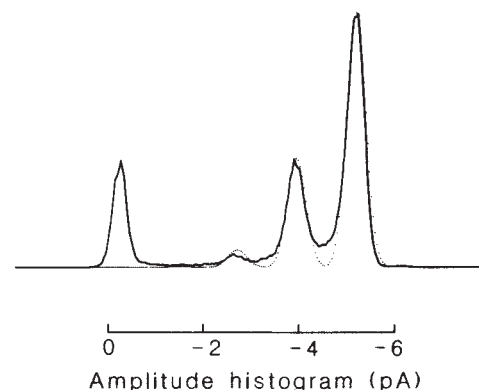


Fig. 2 Amplitude histogram of a 10-s segment of channel activity at a pipette potential of 140 mV. The large peak at the left corresponds to the closed channel current. The three peaks at the right correspond to two, three and four of the subunits being open (the open probability of one subunit is too small to appear in the histogram). Dotted curve, predicted binomial distribution for four independently gating channels, each having a single channel current of 1.25 pA and an open probability of 0.903.

Multi-barrelled K channels in renal tubules

Malcolm Hunter* & Gerhard Giebisch

Department of Physiology, Yale University School of Medicine,
333 Cedar Street, New Haven, Connecticut 06510, USA

Most ion-channels have several kinetically distinct open and closed states. However, a number of groups have reported the presence of sub-conductance states within the open states¹⁻³. It is not clear whether these sub-conductance states represent a partially-occluded open state (with restricted diffusion), or whether they indicate the presence of parallel conductive units. Here we present evidence for the presence of four parallel, equally-conductive subunits that make up a K-selective channel in the apical membrane of early distal tubule in the kidney of *Amphiuma*. The overall open probability of the channel is controlled by a main gate, the kinetics of which can be described by a simple two-state (open or closed) model. This main gate acts either to occlude or to expose the four subunits in unison. The subunits seem to gate independently both of the main gate and of one another, their open probability obeying the binomial distribution.

The diluting segment of amphibians and mammals reabsorbs sodium chloride in excess of water^{4,5}. The uptake of sodium and chloride into the cell from the tubule lumen is through the furosemide-sensitive Na⁺, K⁺, 2Cl⁻-cotransporter. However, the delivery of potassium by the tubule to this segment is not adequate to support the amount of sodium chloride reabsorbed. This apparent dilemma is solved by recycling potassium ions across the apical membrane: the potassium ions that have been transported into the cell diffuse back into the lumen down their electrochemical gradient across the apical membrane by a conductive route⁶. Patch-clamp analysis has identified two types of potassium channel in the apical membranes of these cells⁷. Here we report the gating properties of the small-conductance (30 pS) channel that is probably responsible for the potassium conductance of the apical membrane *in vivo*.

Single early distal tubules were dissected from ventral slices of *Amphiuma* kidney. The tubules were then opened by gentle tearing of the tubule wall with sharpened forceps to expose the apical surfaces of the cells. The tubule was next attached to a holder made from a small coverslip and two strips of aluminium foil, oriented so that the exposed luminal area faced upwards, and transferred to a chamber mounted on the stage of an inverted microscope (for details see ref. 8). Single-channel currents were recorded in the cell-attached mode with 100 mM KCl as the pipette solution. The recordings in this paper use data obtained in three patches with 140 mV potential applied to the pipette (V_p), thus there is a large electrochemical gradient favouring potassium flow from the pipette into the cell.

Figure 1 shows a 200-ms segment of channel activity, with openings in the downward direction (inward current flow). The first impression is that the recording is of a multi-channel patch, but the frequent closures to the zero-current level indicate the presence of a common gating mechanism; we refer to this as

the main gate, which can be thought of as a gate that is capable of completely closing the channel. However, when the main gate is open (when the current is non-zero) the current can be seen to fluctuate between a number of distinct levels. If we assume that there are four parallel subunits in the channel, each with its own gate, then we would expect the currents to be integer multiples of the single subunit current. Evidence that this may be the case is shown in Fig. 1, where the dotted lines are equally spaced at intervals of ~1.25 pA (one quarter of the maximum open-channel current). The open-channel currents can be seen to fluctuate between current levels corresponding to four, three or two of the subunits being open (it is possible to see transitions to the one subunit open condition but these are rare). Changing the pipette potential causes very little change in the channel kinetics but does alter the size of the single-channel current in accordance with Ohm's law ($I = GV$, where I is current, V voltage and G conductance). The relationship between the pipette potential and the channel currents (as deter-

* Present address: Department of Physiology, Leeds University, Leeds LS2 9NQ, UK.

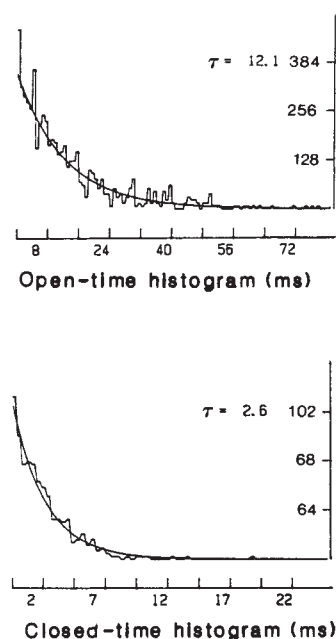


Fig. 3 Open-time (left) and closed-time (right) distributions of the main gate, constructed as described in the text. Curves were fitted by eye, and values of time constant, τ (ms) are also shown.

mined from the peaks of the amplitude histogram—see Fig. 2) gives a subunit slope conductance of 7.9 ± 0.1 pS ($n = 3$), yielding a total single-channel conductance of 31.6 ± 0.5 pS.

Evidence for the existence of the main gate and parallel subunits comes from inspection of the amplitude histogram (Fig. 2). If the histogram were compiled from a patch containing four separate channels, each gating independently of the others but having the same open probability, then the probability density function should be described by the binomial distribution. The very good agreement between the data and the predicted curve indicates that this is the case for the subunits. However, the very large probability of all four of the subunits being simultaneously closed (0.22) cannot be accounted for by the binomial theory (which predicts a probability of 9×10^{-5}), and indicates the presence of an over-riding gating mechanism—the main gate.

The kinetics of the main gate can be derived by measuring the dwell times in either the fully closed (zero current) or open (O_1 , O_2 , O_3 or O_4) states. For a two-state model (closed and open), the rate constant for closing (k_{-1} in Fig. 4) is given by the reciprocal of the mean open time or the time constant of the distribution of open times. Similarly, the rate constant for opening (k_1 of Fig. 4a) is given by the reciprocal of the mean closed time. The open and closed time distributions (Fig. 3) of the main gate can be fitted by single exponentials, indicating that the gating process may be described by this simple two-state model. The open time constant of 12.1 ms corresponds to a rate constant for closing of 82.6 s^{-1} , whereas that of 2.6 ms for the closed times corresponds to a rate constant for opening of 385 s^{-1} . The calculated value for the open probability of the main gate is given by $k_1/(k_1 + k_{-1})$ which is 0.82, exactly equal to the value measured. The mean open-time constant at +140 mV V_p was 10.5 ± 1.1 ms ($n = 3$) and that for closing was 2.1 ± 0.2 ms ($n = 3$). The average measured open probability of the main gate was 0.79 ± 0.01 ($n = 3$) and the estimated open probability was 0.83 ± 0.01 ($n = 3$).

A kinetic evaluation of the rates of transition between the subunits can be obtained by assuming that the kinetic scheme for each subunit comprises two states, closed and open (as for the main gate) and that the rate constants for opening (α) and closing (β) are identical for each channel. The kinetic scheme

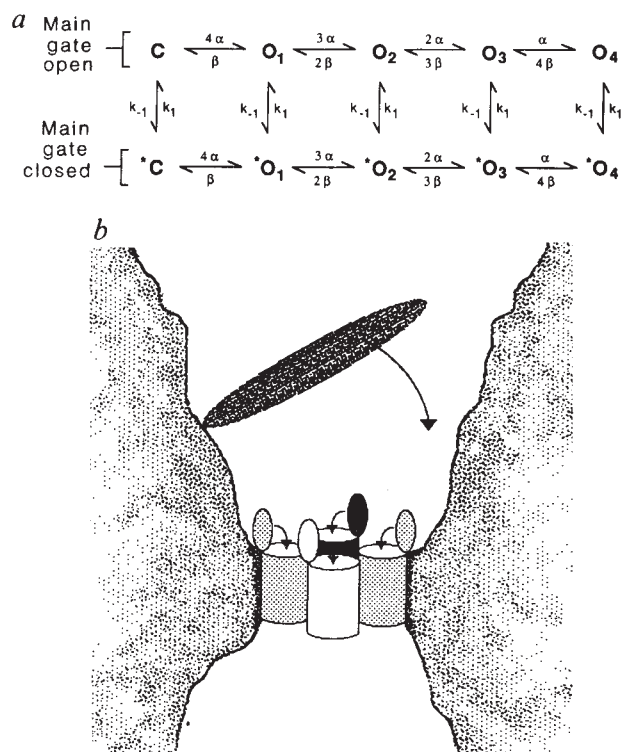


Fig. 4 a, Overall kinetic model of the channel. The states along the top row show the observed transitions between the subunits when the main gate is open. The main gate can close or open (with rates of k_{-1} or k_1 respectively) regardless of the number of open subunits. The states along the bottom row are all closed states (asterisks) caused by the closing of the main gate. Transitions between these closed states can be inferred from the finding that closings of the main gate from, for example O_3 , can be followed by openings to any other state (see text for additional evidence). The estimated values are (in s^{-1}): k_1 , 385; k_{-1} , 83; α , 95 and β , 10.2. b, Diagram illustrating the principal features of the model.

for transitions between sub-conductances is given by either the top or bottom rows of Fig. 4. The dwell time in any one of these states is given by the reciprocal of the sum of the rate constants leaving that state. Thus the lifetime in state O_4 would be $1/(4\beta + k_{-1})$. The distribution of dwell-times in O_4 in the channel in Fig. 1 was also a single exponential having a time constant of 8.1 ms. Solving for β gives a sub-conductance closing rate of 10.2 s^{-1} ; α can be calculated from the relationship between the sub-conductance open probability (estimated from the fit of the binomial distribution) and

$$p_0 = \alpha / (\alpha + \beta)$$

The calculated value for α is 95 s^{-1} .

The lower sequence of sub-conductance states in Fig. 4a (those with a superscript*) indicates that the subunits gate independently not only of each other, but also of the main gate. This can be seen in Fig. 1, where the number of subunits open immediately before and after a transition to the main-gate-closed position is not always the same, because the subunits had gated while the main gate was closed. Qualitative evidence in support of the subunits gating independently of the main gate is that the main gate closings from, or openings to, any sub-conductance level follow the binomial distribution. The probability of opening to levels O_4 , O_3 or O_2 was 0.805, 0.18 and 0.015 respectively and that for closing from these levels was 0.79, 0.19 and 0.026 respectively (344 events). These values closely match each other (indicating that the position of the main gate does not affect the open probability of the sub-units) and the expected distributions for a sub-unit open probability of 0.95.

Multiple sub-conductance states have been observed in other ion channels. For instance, Miller¹ suggested a scheme similar to that proposed here for the chloride selective channel of *Torpedo* electroplax, with two sub-units. Gage and co-workers³ proposed that the large anion-selective channel of alveolar epithelial cells was composed of six 'co-channels' which were in turn regulated by a common gating mechanism. Miller had speculated that the property of channels with discrete sub-conductances may be restricted to anion-selective channels, but these studies and others⁹ demonstrate that this is not the case.

What is the advantage of this type of channel structure? It has been suggested that multiple subunits in a main channel provide the channel with a high conductance while maintaining high selectivity³. An alternative idea is that it allows for more varied regulation of the channel; for example, the subunit gates may be sensitive to the intracellular pH whereas the main gate could be susceptible to phosphorylation. We suggest implications for 'normal' channels, which frequently show many components in their dwell-time distributions. Such kinetic behaviour may not be due to variations in the gating kinetics of an individual gate but instead may indicate that a number of gates are arranged in series along the channel.

This work was supported by an NIH grant. We thank Dr F. Sigworth for helpful suggestions, Ms Victoria Guckian and Amabel Shih for preparing Fig. 4 and Drs D. Lipscombe and C. Benham for comments on the manuscript.

Received 3 March; accepted 14 April 1987.

1. Miller, C. *Phil. Trans. R. Soc. B299*, 401-411 (1982).
2. Geletyuk, V. I. & Kazachenko, V. N. *J. Membrane Biol.* **86**, 9-15 (1985).
3. Krouse, M. E., Schneider, G. T. & Gage, P. W. *Nature* **319**, 58-60 (1986).
4. Oberleithner, H., Giebisch, G., Lang, F. & Wang, W. *Pflügers Arch. ges. Physiol.* **60**, 1173-1179 (1982).
5. Greger, R. & Schlatter, E. *Pflügers Arch. ges. Physiol.* **396**, 325-334 (1983).
6. Greger, R. & Schlatter, E. *Pflügers Arch. ges. Physiol.* **392**, 92-94 (1981).
7. Hunter, M., Kawahara, K. & Giebisch, G. *Fedn. Proc.* **45**, 2723-2726 (1986).
8. Hunter, M., Lopes, A. J., Boulpaep, E. L. & Giebisch, G. *Am. J. Physiol.* **251**, F725-F733 (1986).
9. Kazachenko, V. N. & Geletyuk, V. I. *Biochim. biophys. Acta* **773**, 132-142 (1984).

Nitric oxide release accounts for the biological activity of endothelium-derived relaxing factor

R. M. J. Palmer, A. G. Ferrige & S. Moncada*

Wellcome Research Laboratories, Langley Court, Beckenham, Kent BR3 3BS, UK

Endothelium-derived relaxing factor (EDRF) is a labile humoral agent which mediates the action of some vasodilators. Nitrovasodilators, which may act by releasing nitric oxide (NO), mimic the effect of EDRF and it has recently been suggested by Furchgott¹ that EDRF may be NO. We have examined this suggestion by studying the release of EDRF and NO from endothelial cells in culture. NO was determined as the chemiluminescent product of its reaction with ozone². The biological activity of EDRF and of NO was measured by bioassay³. The relaxation of the bioassay tissues induced by EDRF was indistinguishable from that induced by NO. Both substances were equally unstable. Bradykinin caused concentration-dependent release of NO from the cells in amounts sufficient to account for the biological activity of EDRF. The relaxations induced by EDRF and NO were inhibited by haemoglobin and enhanced by superoxide dismutase to a similar degree. Thus NO released from endothelial cells is indistinguishable from EDRF in terms of biological activity, stability, and susceptibility to an inhibitor and to a potentiator. We suggest that EDRF and NO are identical.

* To whom correspondence should be addressed.

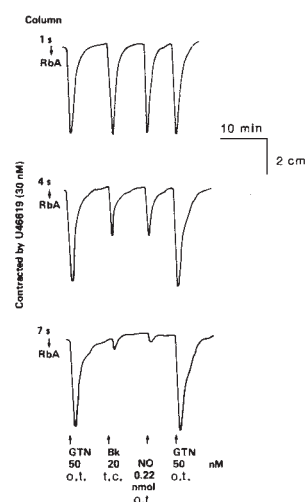


Fig. 1 Relaxation of rabbit aortae by EDRF and NO. A column packed with endothelial cells cultured on microcarriers was perfused with Krebs' buffer (5 ml min⁻¹). The effluent was used to superfuse three spiral strips of rabbit aorta (RbA), denuded of endothelium, in a cascade. The tissues were contracted sub-maximally by a continuous infusion of 9,11-dideoxy-9 α , 11 α -methano epoxy-prostaglandin F₂ α (U46619; 30 nM) and were separated from the cells by delays of 1, 4 and 7 s, respectively. The sensitivity of the bioassay tissues was standardized by administration of glyceryl trinitrate (GTN; 50 nM) over the tissues (o.t.). EDRF was released from the cells by a 1 min infusion through the column (t.c.) of bradykinin (Bk, 20 nM). NO (0.22 nmol) was dissolved in He-deoxygenated H₂O and administered as a 1 min infusion.

EDRF was released from porcine aortic endothelial cells ($2-5 \times 10^7$ cells) cultured on microcarriers and was detected by bioassay on spiral strips of rabbit aorta superfused in a cascade as described previously³. In some experiments the endothelial cells were replaced by porcine aortic smooth muscle cells or 3T3 cells ($1.5-1.9$ and $2.4-3.9 \times 10^7$ cells respectively). A gas bulb was filled with NO (British Oxygen) from a cylinder and sealed with silicone rubber injection septa. An appropriate volume (10-1,000 μ l) was removed with a syringe and injected into another gas bulb filled with H₂O, which had been deoxygenated by gassing with He for 1 h, to give stock solutions of NO of 0.01-1.0% (v/v). These solutions were stable for ~ 10 h. NO was determined by chemiluminescence².

The bioassay tissues were relaxed by glyceryl trinitrate (GTN, 50 nM), by EDRF released from the cells by bradykinin (20 nM) and by authentic NO (0.22 nmol; Fig. 1; $n = 30$). The relaxations induced by EDRF and NO declined during passage down the cascade at rates which were not significantly different from each other, and were 50% of that of the uppermost bioassay tissue (the 'half-life') after 3.6 ± 0.1 and 4.1 ± 0.2 s (mean $\pm$ s.e.m., $n = 4$) respectively. In contrast, the relaxations caused by GTN did not decline during passage down the cascade at any of the concentrations used (1-50 nM).

The stability of EDRF or of authentic NO in Krebs' buffer was determined using bioassay by interposing lengths of tubing, between either the generator cells or the point of infusion of NO (0.22 nmol) and the uppermost detector tissue, so as to vary the transit time. The half-life in transit of EDRF was not significantly different from that of NO (30.8 ± 1.9 and 30.4 ± 2.2 s; $n = 3$ respectively). The fact that both EDRF and NO disappear more rapidly during passage down the cascade than they do in Krebs' buffer alone may indicate that the tissues contribute in some way to their decay, either by generating superoxide anions (O_2^-) or by some unknown mechanism. Alternatively, the inter-

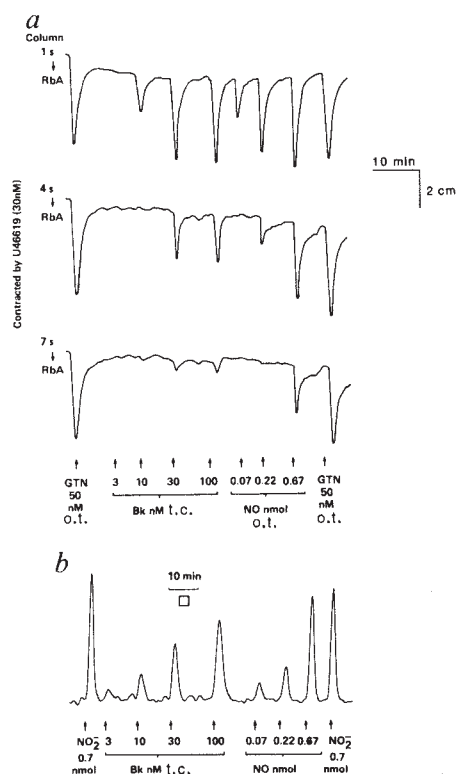


Fig. 2 *a*, Bioassay. Relaxation of rabbit aortae by EDRF and NO. The bioassay tissues were relaxed in a concentration-dependent manner by EDRF released from the cells by bradykinin (Bk; 3–100 nM t.c.) and by NO (0.07–0.67 nmol, o.t.) as in Fig. 1. *b*, Chemiluminescence. Release of NO by bradykinin (Bk) from a replicate column of the cells used in the bioassay. The amounts of NO (administered as 1 min infusion into the column effluent) which relaxed the bioassay tissues were also detectable by chemiluminescence. Effluent from the column, or Krebs' buffer into which authentic NO was injected, was passed continuously (5 ml min⁻¹) into a reaction vessel containing 75 ml 1.0% sodium iodide in glacial acetic acid under reflux. NO was removed from the refluxing mixture under reduced pressure in a stream of N₂, mixed with ozone and the chemiluminescent product measured with a photomultiplier. The amounts of NO detected were quantified after correcting for baseline drift using a polynomial fit and reducing electrical noise, by Fourier transformation and application of a Gaussian function. The areas under the peaks were converted to nmol of NO by reference to a NO₂⁻ standard curve. Similar results were obtained in two other experiments. □, Area equivalent to 0.22 nmol NO.

action of NO and O₂ in the delay tubes may differ from that during passage down the cascade.

Bradykinin released EDRF from all the endothelial cell cultures tested ($n > 30$). At concentrations ranging from 3 nM to 100 nM the release of EDRF was concentration-dependent and caused relaxations of the bioassay tissues similar to those induced by NO (0.07–0.67 nmol; Fig. 2*a*). Bradykinin (3–100 nM) also caused concentration-dependent release of amounts of NO sufficient to account for the relaxations of the bioassay tissues induced by EDRF (Fig. 2*b*). The release induced by 3, 10, 30 and 100 nM bradykinin was 0.09 ± 0.02 , 0.19 ± 0.01 , 0.44 ± 0.02 , and 0.68 ± 0.02 nmol respectively ($n = 3$). Furthermore, there was a correlation ($r = 0.94$, $n = 12$) between the amount of NO released and the relaxation of the bioassay tissues induced by EDRF. Release of EDRF or NO by bradykinin (100 nM) could not be detected (< 0.05 nmol) from porcine aortic smooth muscle cells (2 experiments) or 3T3 cells (2 experiments).

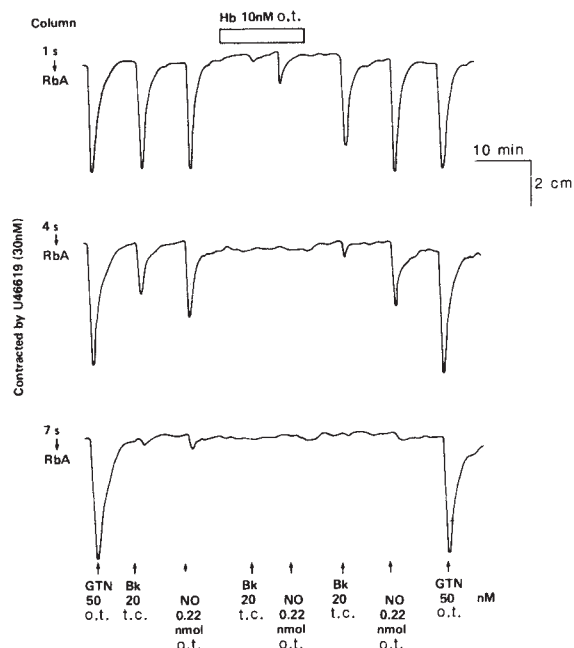


Fig. 3 The effect of purified human haemoglobin (Hb; $< 3\%$ Met haemoglobin) on the relaxation of rabbit aortae induced by EDRF and NO. Effluent of a column containing endothelial cells was used to superfuse three spiral strips of rabbit aorta in a cascade (as in Fig. 1). The bioassay tissues were relaxed by NO (0.22 nmol o.t.) or by EDRF released by bradykinin (Bk, 20 nM t.c.) as in Fig. 1. Infusion of Hb (10 nM o.t.) inhibited the relaxations induced by both EDRF and NO to a similar extent. Larger concentrations (> 100 nM) of Hb were needed to inhibit the relaxations induced by GTN (data not shown) confirming a previous report⁷. Similar results were obtained in two other experiments.

NO reacts readily with O₂ to produce NO₂, which then forms NO₂⁻ and NO₃⁻ in neutral aqueous solution according to the following reactions:



Neither NO₂⁻ nor NO₃⁻ (as their sodium salts) cause relaxation of the vascular strips at concentrations below 10 μM (data not shown). These reactions therefore represent an inactivation mechanism for NO in terms of the biological activity we measure and thus the decay of NO can be studied by bioassay.

There was a delay of 1 s between the endothelial cells, or the point of infusion of authentic NO, and the reflux vessel from which NO was recovered before measurement. Because the half-life of NO and EDRF determined by bioassay is ~30 s, little decay of NO to NO₂⁻ would occur in this time. Therefore we detect NO directly rather than NO derived from the reduction of NO₂⁻ in the reflux vessel. This is supported by the finding that the chemiluminescence detected from 0.7 nmol NO₂⁻ and from 0.67 nmol authentic NO are similar (Fig. 2*b*). If NO decayed completely then a maximum of 50% of the NO would be detected, for NO₃⁻ is not reduced to NO in the reflux vessel. Thus the chemiluminescence method cannot be used to determine the decay of NO when longer delays are used.

NO₂⁻ is not converted to NO when refluxing glacial acetic acid (without sodium iodide) is diluted $> 50\%$ with Krebs' buffer. Under these conditions, NO released from the endothelial cells could be detected. Furthermore, the decay of NO could also be determined although larger amounts were required (2.25 nmol). The decay was found to approximate to second-order kinetics (data not shown), consistent with the reaction (1) above. Second-order decay kinetics may help to explain the variations in half-life reported for EDRF^{4,5}.

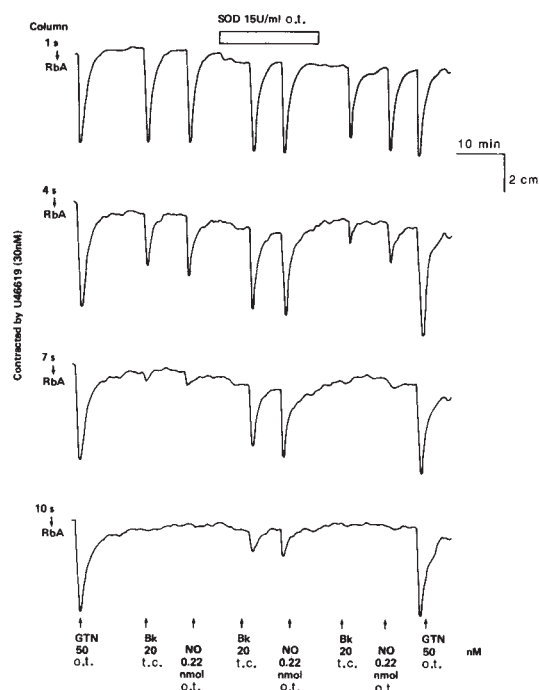


Fig. 4 The effect of SOD on the relaxation of rabbit aortae induced by EDRF and NO. Effluent of a column of endothelial cells was used to superfuse four strips of rabbit aorta in a cascade (as in Fig. 1). The bioassay tissues were relaxed by NO (0.22 nmol, o.i.) or by EDRF released by bradykinin (20 nM Bk t.c.). Infusion of SOD (15 U ml⁻¹ o.i.) increased the stability of EDRF and NO to a similar extent, so that the fourth tissue was relaxed by both compounds. Similar results were obtained in two other experiments.

Haemoglobin, which has been reported to bind EDRF^{6,7} and NO⁸, inhibited the relaxations of the bioassay tissues induced by EDRF and NO (Fig. 3), with an IC₅₀ of 3.6 ± 0.6 and 8.1 ± 1.4 nM (*n* = 3) respectively. Superoxide dismutase (SOD) reduces the inactivation of EDRF^{9,10}. Infusions of SOD (15 U ml⁻¹) over the tissues reduced the decay of both EDRF and authentic NO to a similar extent (Fig. 4), indicating that O₂⁻ contributes to the inactivation of both compounds.

NO is therefore released from vascular endothelial cells, but not from smooth muscle cells or 3T3 cells, in amounts sufficient to account for the relaxations attributed to EDRF. Furthermore, it has the same biological activity and chemical stability as EDRF and its action is as susceptible to inhibition by haemoglobin and to potentiation by SOD. Thus, the biological activity of EDRF is accounted for by NO when considered in terms of the classical criteria proposed by Dale¹¹ for the identification of a mediator. The proposal that EDRF is NO is consistent with our suggestion that EDRF is inactivated by O₂⁻ generated in the oxygenated Krebs' buffer⁹ or by the action of some inhibitors of EDRF¹², for O₂⁻ reacts with NO to form NO₃⁻ ultimately¹³. Whether NO is released from vascular endothelial cells as NO itself, or as a precursor which gives rise to NO following release, deserves further investigation.

We thank Dr C. L. Walters for advice and assistance with the chemiluminescence method, N. A. Foxwell, C. J. Saxby, B. C. Sweatman, Ms L. J. Wallis and M. J. Ashton for technical assistance, and our colleagues at the Wellcome Research Laboratories for helpful discussions.

Received 28 January; accepted 10 April 1987.

- Griffith, T. M., Edwards, D. H., Lewis, M. J., Newby, A. C. & Henderson, A. H. *Nature* **308**, 645-647 (1984).
- Forstermann, U., Trogisch, G. & Busse, R. *Eur. J. Pharmac.* **106**, 639-643 (1985).
- Cocks, T. M. & Angus, J. A. in *Vascular Neuroeffector Mechanisms* (eds Bevan, J. A., Godfraind, T., Maxwell, R. A., Stoclet, J. C. & Worcel, M.) 131-136 (Elsevier, Amsterdam, 1985).
- Martin, W., Villani, G. M., Jothianandan, D. & Furchgott, R. F. *J. Pharmac. exp. Ther.* **232**, 708-716 (1985).
- Gibson, Q. H. & Roughton, F. J. W. *J. Physiol., Lond.* **136**, 507-526 (1957).
- Gryglewski, R. J., Palmer, R. M. J. & Moncada, S. *Nature* **320**, 454-456 (1986).
- Rubanyi, G. M. & Vanhoutte, P. M. *Am. J. Physiol.* **250**, H222-H227 (1986).
- Dale, H. H. *Bull. Johns Hopkins Hosp.* **53**, 297-347 (1933).
- Moncada, S., Palmer, R. M. J. & Gryglewski, R. J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 9164-9168 (1986).
- Blough, N. V. & Zafiriou, O. C. *Inorg. Chem.* **24**, 3502-3504 (1985).

Different recombination site specificity of two developmentally regulated genome rearrangements

James W. Golden*, Martin E. Mulligan & Robert Haselkorn

Department of Molecular Genetics and Cell Biology, The University of Chicago, Chicago, Illinois 60637, USA

In the absence of a combined nitrogen source, such as ammonia, approximately every tenth vegetative cell along filaments of the cyanobacterium *Anabaena* develops into a heterocyst, a terminally differentiated cell that is morphologically and biochemically specialized for nitrogen fixation¹. At least two specific DNA rearrangements involving the nitrogen-fixation (*nif*) genes occur during heterocyst differentiation², one within the *nifD* gene and the other near the *nifS* gene. The two rearrangements have several properties in common. Both occur quantitatively in all heterocyst genomes, both occur at approximately the same developmental time, late in the process of heterocyst differentiation², and both result from site-specific recombination between short repeated DNA sequences. We report here the nucleotide sequences found at the site of recombination near the *nifS* gene. These sequences differ from those found previously for the *nifD* rearrangement², suggesting that the two rearrangements are catalysed by different enzymes and may be regulated independently. We also show that the *nifS* gene is transcribed only from rearranged genomes.

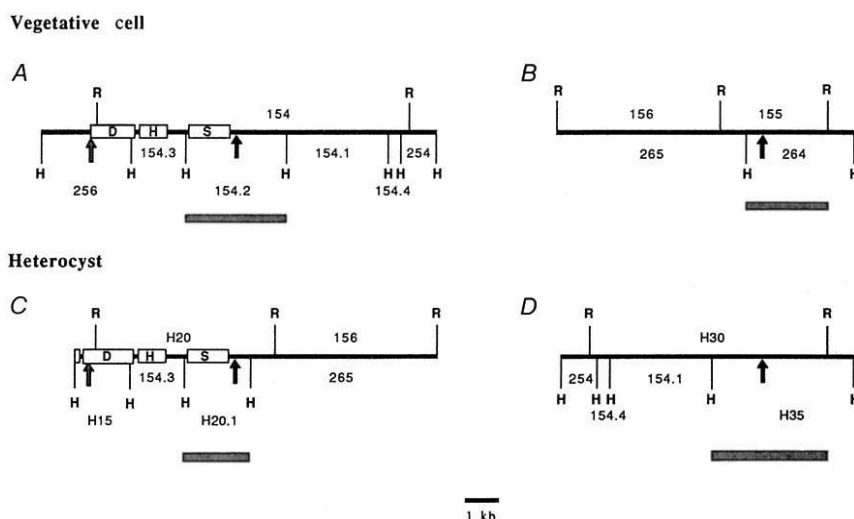
The *nifD* rearrangement is the deletion of an 11-kilobase (kb) element, the *nifD* excision, from just inside the 3' end of the *nifD* gene². The *nifD* excision persists in heterocysts as a circular molecule, but it has no known function. The excision is the result of site-specific recombination between 11-base pair (bp) directly repeated sequences found at the ends of the *nifD* excision. Excision restores the complete *nifD* open reading frame³ and forms an operon containing the *nifH*, *nifD* and *nifK* genes, which encode the three polypeptide components of the nitrogenase complex. The gene *xisA*, located near the *nifK* proximal end of the *nifD* excision, is necessary for the rearrangement of *Anabaena nifHDK* DNA in *Escherichia coli*, and presumably encodes the recombinase responsible for the *nifD* rearrangement⁴.

The second known rearrangement of the vegetative chromosome in *Anabaena* involves DNA located near the *nifS* gene. This gene was previously designated *nifV/S*, ref. 5. However, Southern hybridization experiments using this gene as a probe against restriction digests of *Klebsiella* DNA show that the gene is *nifS* (M.E.M., unpublished). The *nifS* gene is required for an unknown step in the maturation of the nitrogenase enzyme. The *nifS* rearrangement was first identified when a Southern blot of *EcoRI*-digested genomic DNA from vegetative cells and from purified heterocysts was probed with the *HindIII* fragment, An154.3, containing the *nifH* gene². The *nifH* gene was found

* Present address: Department of Biology, Texas A & M University, College Station, Texas 77843, USA.

- Furchgott, R. F. in *Mechanisms of Vasodilatation* Vol. IV (ed. Vanhoutte, P. M.) (Raven, New York, in the press).
- Downes, M. J., Edwards, M. W., Elsey, T. S. & Walters, C. L. *Analyst* **101**, 742-748 (1976).
- Gryglewski, R. J., Moncada, S. & Palmer, R. M. J. *Br. J. Pharmac.* **87**, 685-694 (1986).

Fig. 1 Restriction maps of the *nifS* rearrangement DNA breakpoints in the vegetative cell and heterocyst genomes. Restriction sites are labelled R (*EcoRI*) and H (*HindIII*). Restriction fragment names are indicated without the 'An' prefix. Restriction fragments common both to vegetative cells and to heterocysts have the same numerical designation. We have renamed only those heterocyst restriction fragments that contain DNA breakpoints and have indicated that they are unique to the heterocyst genome by an 'H' in their designation. The *nifS* rearrangement breakpoints are shown by the solid arrows. In the vegetative cell they are located in the *HindIII* restriction fragments, A, An154.2 and B, An264. In the heterocyst they are located in the *HindIII* restriction fragments, C, AnH20.1 and D, AnH35. Two of the *nifD* rearrangement DNA breakpoints are shown by the shaded arrows. They occur in, A, An256 and C, AnH15. The locations of the *nifD* and *nifH* genes and the approximate location of the *nifS* gene are shown; all three genes are transcribed from right to left. The *nifD* gene is incomplete in An256. The 5' end of the *nifK* gene is also indicated on AnH15 but is not labelled. The hybridization probes used in Fig. 2 are shown as shaded bars below each of the restriction maps.



on a 10.5-kb *EcoRI* fragment in vegetative cells but this fragment was replaced by a 6.0-kb *EcoRI* fragment in the heterocyst genome. A DNA breakpoint was located in the 3.3-kb *HindIII* fragment An154.2 close to the 5' end of the *nifS* gene. When An154.2 was used to probe *HindIII*-digested genomic DNA, a self-complementary band of 3.3 kb was seen in vegetative cell DNA, but in heterocyst DNA this band was replaced by two new bands at 4.7 and 2.3 kb (ref. 2).

If a recombination event similar to the one that occurs in the *nifD* gene is responsible for the *nifS* rearrangement, then a second DNA breakpoint should be present in the vegetative cell genome, and the heterocyst genome should contain the two DNA breakpoints representing the products of rearrangement. The remaining rearrangement breakpoints were isolated from vegetative cell and heterocyst genomic DNA libraries as follows. The 6-kb *EcoRI* fragment in the heterocyst genome containing the *nifS* gene, AnH20 (Fig. 1C), was isolated from a λ gt7 (ref. 6) library of heterocyst *EcoRI* fragments using An154.3 as a probe. It was subcloned into the plasmid vector pBR328 (ref. 7). The 7.9-kb *EcoRI* fragment AnH30 (Fig. 1D) was isolated from the λ gt7 heterocyst DNA library using as a probe a small *HincII* fragment internal to the *HindIII* fragment An154.2 and between the DNA breakpoint and An154.1 (Fig. 1A). The final DNA breakpoint, in the *EcoRI* fragment An155 in the vegetative cell chromosome (Fig. 1B), was obtained on two different overlapping clones isolated from a λ L47.1 library⁸ of *Sau3AI* partially digested vegetative cell DNA cloned into the *Bam*HI site of the vector. The probe used to isolate these clones was the *HindIII*-*EcoRI* fragment that is common to AnH20 and An265 (Fig. 1C). Both the AnH30 and An155 *EcoRI* fragments were subcloned into pUC18 (ref. 9).

Figure 1 shows the restriction map of the chromosome around each of the four DNA breakpoint clones, two from the vegetative cell chromosome and two from the heterocyst chromosome after rearrangement. The restriction maps and Southern analysis of the breakpoint clones indicate that the rearrangement is the result of a recombination event producing a reciprocal exchange between two sites. The recombination results in the juxtaposition of the left portion of An154.2 containing the *nifS* gene with DNA from An264 to form the new *HindIII* fragment, AnH20.1, found in the heterocyst chromosome. The reciprocal product of the rearrangement is the *HindIII* fragment AnH35 which includes the right half of An154.2 (adjacent to An154.1) and the majority of An264 including the *EcoRI* site (Fig. 1D).

Southern blot analysis of genomic DNA confirmed that the breakpoint clones represented the *in vivo* *nifS* rearrangement

substrates and products, and were not artefacts of cloning. Figure 2 shows *HindIII* digested genomic DNA, isolated both from vegetative cells and from purified heterocysts, probed with each of the four DNA breakpoints. In each instance the probe hybridized only with itself or with the expected rearranged fragments. The vegetative cell breakpoint probes, An154.2 and An264, hybridized with the expected single bands in vegetative cell DNA (3.3 and 3.7 kb respectively), and as predicted by a reciprocal recombination event, hybridized with the same two

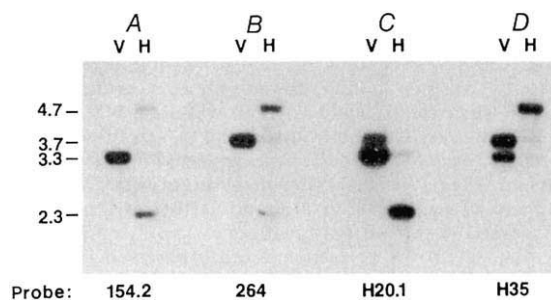


Fig. 2 Southern analysis of the restriction fragments containing the four *nifS* DNA breakpoints in vegetative cells and heterocysts. *HindIII*-digested DNA from vegetative cells (V) and from purified heterocysts (H) was hybridized with the following probes: An154.2 (A), the large *HindIII*-*EcoRI* fragment from An264 (B), AnH20.1 (C), the large *HindIII*-*EcoRI* fragment from AnH35 (D) see Fig. 1. Sizes are in kb. The two vegetative cell probes hybridized with the same two bands in heterocyst DNA digests (2.3 and 4.7 kb). The two heterocyst probes hybridized with the same two bands in vegetative cell DNA digests (3.3 and 3.7 kb). The heterocyst DNA preparation contains a trace amount of vegetative cell DNA and so faint bands are observed corresponding to the vegetative cell fragments. The relative intensities of the two bands at 4.7 and 2.3 kb in the heterocyst DNA lanes of A and B are consistent with the relative contributions of the probe. Similarly, the relative intensities of the two bands at 3.7 and 3.3 kb in the vegetative DNA lanes of C and D are consistent with the relative contributions of the probe.

Methods. Heterocysts were purified as described by Golden *et al.*² DNA isolated from vegetative cells or from purified heterocysts ($\approx 1 \mu\text{g}$) was digested with *HindIII* and electrophoresed through a 0.7% agarose gel. The gel was blotted onto nitrocellulose and probed as indicated following standard procedures¹². The fragments used as probes were cloned into pBR322 (An154.2) or pUC18 (AnH20.1 and the large *HindIII*-*EcoRI* fragments of An264 and AnH35). Fragments were isolated from agarose gels¹³ and nick-translated using standard procedures¹².

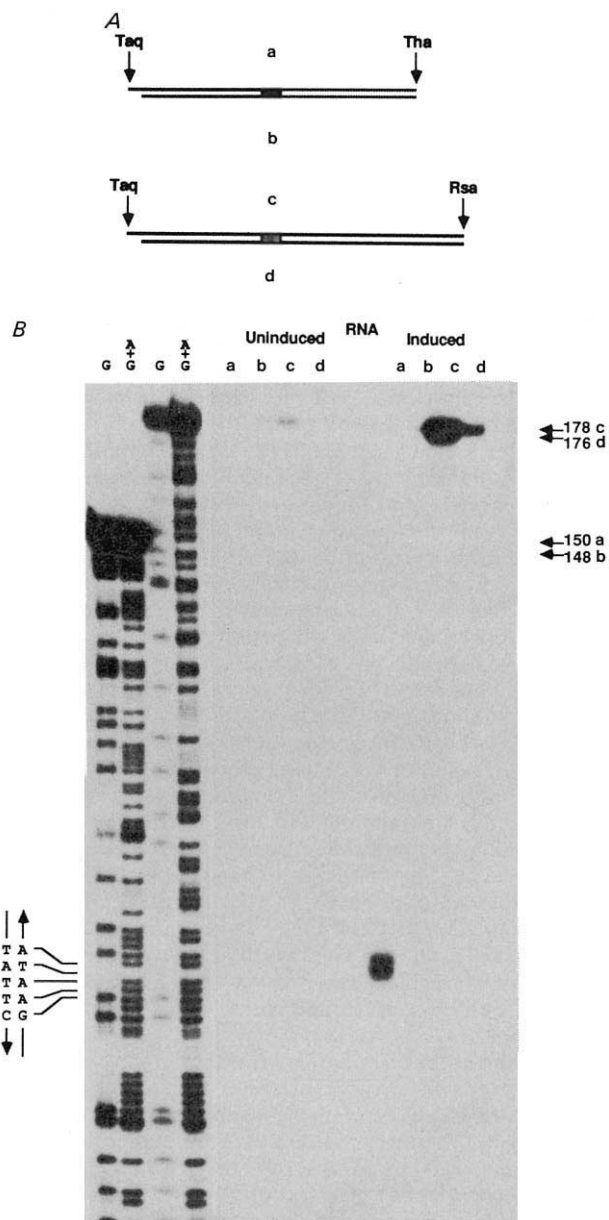
Fig. 3 *A*, DNA sequence of the *nifS* rearrangement breakpoints and *B*, comparison of the conserved sequences involved in site-specific recombination for the *nifS* and *nifD* genome rearrangements. The sequence of the An154.2 and AnH20.1 breakpoints is shown 5' to 3' on the *nifS* noncoding DNA strand; the corresponding strands of An264 and AnH35 are shown. The central sequence within which recombination occurs is underlined for each of the individual *nifS* rearrangement breakpoint sequences and in the consensus sequences for both the *nifS* and *nifD* rearrangements. A more extended *nifS* consensus sequence is indicated by the dashed underline. The *nifS* consensus sequence is derived from the sequence in *A* and the *nifD* consensus sequence is from Golden *et al.*².

Methods. DNA fragments containing each of the four breakpoints were isolated from agarose gels¹³, end-labelled¹², strand-separated and sequenced by the method of Maxam and Gilbert¹⁴. The DNA sequence was determined on both strands for the An154.2 and AnH20.1 breakpoints and on one strand for the An264 and AnH35 breakpoints.

Vegetative cell breakpoints	An154.2	ATAGCAGATACTTC <u>TATTC</u> CAGAAATTTCCGAACAG
	An264	GTTGGTTTAATTGT <u>TATTC</u> CAGAGAAGTTTACACTATCA
Heterocyst breakpoints	AnH20.1	GTTGGTTTAATTGT <u>TATTC</u> CAGAAATTTCCGAACAG
	AnH35	ATAGCAGATACTTC <u>TATTC</u> CAGAGAAGTTTACACTATCA
Consensus sequences	<i>nifS</i>	-T-G-----A-T-- <u>TATTC</u> --AGAA-TTT-C---A---
	<i>nifD</i>	--AA--CAC-----GGATTACTCCG--C-T---ACGG

Fig. 4 *S*₁ analysis of the effect of rearrangement on transcription of *nifS*. *A*, Map of the DNA fragments used as probes. A DNA fragment including the breakpoint of the rearrangement was isolated from An154.2 using the restriction endonucleases *Taq*I and *Tha*I. The fragment was end-labelled and the individual strands separated into a 150-base fragment labelled at the *Taq*I end (a) and a 148-base fragment labelled at the *Tha*I end (b). The corresponding fragment was isolated from AnH20.1 using *Taq*I and *Rsa*I and the separated end-labelled fragments are 178 (c) and 176 (d) bases long respectively. The two double-stranded fragments are identical from the *Taq*I site up to and including the breakpoints (shaded) of the rearrangement. *B*, Each of the separate end-labelled strands was hybridized to RNA isolated from uninduced (N+) and induced (N-) cultures of *Anabaena*, and digested with nuclease *S*₁. All four strands showed only background levels of protection when uninduced (N+) RNA was used. When induced (N-) RNA was used, the heterocyst strand (c) was fully protected whereas the vegetative strand (a) was protected only as far as the breakpoint. The full-length protection observed with d was due to the presence of a small amount of strand c. Otherwise, the opposite strands b and d showed only background levels of protection. Background levels of *S*₁ protection were also observed when tRNA was used (not shown). DNA sequencing ladders (G and G+A reactions) of the 150- and 178- base strands are shown and the 5-bp conserved sequence of the breakpoint is indicated.

Methods. DNA fragments containing the breakpoints were isolated from agarose gels¹³, end-labelled¹² and strand-separated on a 7 M urea/6% acrylamide-sequencing gel. RNA was isolated from *Anabaena* as described previously². Nuclease *S*₁ digestions, for each of which 200 µg of RNA was used, were performed as described¹⁵ except that hybridization was for 4 h at 34 °C and *S*₁ nuclease digestion was at 37 °C. The products of digestion were electrophoresed on a 7 M urea/6% acrylamide-sequencing gel. G and A+G sequencing reactions were carried out on the 150- and 178-base strands using the method of Maxam and Gilbert¹⁴.



bands of 2.3 kb and 4.7 kb in heterocyst DNA, which are the fragments AnH20.1 and AnH35, respectively (Fig. 2A and B). The probes obtained from heterocyst clones, AnH20.1 and AnH35, hybridized with single bands in heterocyst DNA (2.3 and 4.7 kb respectively) that are replaced in vegetative cell DNA by bands corresponding to An154.2 and An264 at 3.3 kb and

3.7 kb (Fig. 2C and D). All these clones represent unique sequences in the *Anabaena* genome.

The DNA breakpoints in An154.2 and AnH20.1 were localized to small restriction fragments so that the DNA sequence of this region could be determined. Each of the purified *Hind*III fragments, An154.2 and AnH20.1, was digested with restriction

enzymes that have four-base recognition sequences and the resultant fragments were separated in adjacent lanes of an agarose gel. A Southern blot of the gel was then probed with labelled An154.2. The restriction fragments of AnH20.1 that were completely derived from An264 did not hybridize with the probe, whereas those that were completely derived from An154.2 hybridized with the probe and corresponded directly to bands in the adjacent lane containing digested An154.2. However, a single band in the AnH20.1 digest that hybridized with the An154.2 probe did not have a corresponding band in the An154.2 lane. This fragment contained the *nifS* rearrangement breakpoint. The same procedure was used for the other breakpoint clones to identify small restriction fragments which were suitable for sequencing.

The DNA sequences surrounding each of the four breakpoints are shown in Fig. 3A, oriented 5' to 3' relative to the *nifS* noncoding DNA strand. The DNA sequence shows that the *nifS* rearrangement is a site-specific recombination between short repeated elements. A reciprocal exchange occurs between the two vegetative cell sequences found in An154.2 and An264 to produce the two new fragments in the heterocyst genome, AnH20.1 and AnH35. The rearrangement is conservative in that there is neither a loss nor gain of any DNA sequence. The recombinational cross-over must occur within the five base-pair sequence TATTC, although the recognition sequence presumably includes part or all of the adjacent conserved sequence shown in the figure.

Although the *nifS* and *nifD* rearrangements have several properties in common, there is no homology between the conserved sequences found at the site of recombination (Fig. 3B). The difference in the site specificity between the two rearrangements is perhaps not surprising because of their close proximity on the chromosome. If they shared the same sites it might be difficult to avoid illegitimate rearrangements. The difference in recombination site sequences, which presumably reflects a difference in the enzymes that catalyse the two rearrangements, allows for differential regulation of the two rearrangements. Although the *nifS* and *nifD* rearrangements show identical temporal control under normal conditions of heterocyst development, they can be uncoupled under some conditions (M.E.M. and G. J. Schneider, unpublished data).

The *nifD* rearrangement is a deletion of 11 kb from the chromosome resulting from recombination between 11-bp direct repeats found at the ends of the 11-kb element. The *nifS* rearrangement also results from recombination between repeated DNA sequences. Preliminary chromosome walking experiments, using cosmid libraries of *Anabaena* DNA, suggest that the two breakpoints are relatively distant in vegetative cell DNA, probably more than 50 kb apart (J.W.G., unpublished). The rearrangement is probably the deletion of a circular molecule.

The *nifD* rearrangement has two specific effects on the expression of the *nif* genes: one is to form the complete *nifD* coding sequence, and the second is to place the *nifK* gene under the control of the *nifH* promoter. Much less is known about the effect of the *nifS* rearrangement on gene expression. The DNA breakpoint in An154.2 is very close to the 5' end of the coding sequence for the *nifS* gene (M.E.M., unpublished). DNA sequence analysis of the region that is joined to *nifS* as a result of the rearrangement shows considerable homology with *nifB* (refs 10, 11 and M.E.M., unpublished). Therefore the *nifB* gene is juxtaposed with *nifS* as a result of the rearrangement. It is likely that the expression of the *nifS* gene is affected by the rearrangement. Transcription of *nifS* occurs only after rearrangement has taken place. We isolated end-labelled single-stranded DNA fragments spanning the breakpoints in An154.2 and AnH20.1. These fragments were used to probe RNA that was isolated from cultures of *Anabaena* grown in the presence (uninduced) and in the absence (induced) of a fixed nitrogen source. The results are shown in Fig. 4. Comparison of the results obtained using RNA from an uninduced culture with

those obtained using RNA from an induced culture shows that transcription of this region occurs only under inducing conditions. Furthermore, comparison of the results obtained using a vegetative cell probe (An154.2) with those obtained using a heterocyst probe (AnH20.1) show that transcription occurs only after rearrangement has taken place. The vegetative cell probe was protected from *S*₁ nuclease digestion only to the extent that its sequence is identical to that of the heterocyst probe, that is only from the labelled end up to and including the breakpoint sequence. By contrast, the heterocyst probe was fully protected. These results suggest that rearrangement is a prerequisite for transcription of *nifS*. We do not know, at present, whether the rearrangement has any effect on gene structure or expression in the other breakpoint fragments, An264 and AnH35. As is the case with the *nifD* rearrangement, we do not know why *Anabaena* has adopted this unique method of gene and developmental regulation.

We thank G. Schneider for unpublished data and W. Buikema for critically reading the manuscript. This work was supported by research grants from the NIH, the USDA/SEA through the Competitive Research Grants Office, and the Monsanto Company. J.W.G. was supported by an NIH postdoctoral fellowship.

Received 12 March; accepted 22 April 1987.

- Haselkorn, R. *Rev. Pl. Physiol.* **29**, 319-344 (1978).
- Golden, J. W., Robinson, S. J. & Haselkorn, R. *Nature* **314**, 419-423 (1985).
- Haselkorn, R., Golden, J. W., Lammers, P. J. & Mulligan, M. E. in *Nitrogen Fixation Research Progress* (eds Evans, H. J., Bottomly, P. J. & Newton, W. E.) 485-490 (Nijhoff, Dordrecht, 1985).
- Lammers, P. J., Golden, J. W. & Haselkorn, R. *Cell* **44**, 905-911 (1986).
- Rice, D., Mazur, B. J. & Haselkorn, R. *J. biol. Chem.* **257**, 13157-13163 (1982).
- Davis, R. W., Botstein, D. & Roth, J. R. *Advanced Bacterial Genetics* (Cold Spring Harbor Laboratory, New York, 1980).
- Soberon, X., Covarrubias, L. & Bolivar, F. *Gene* **9**, 287-305 (1980).
- Loenen, W. A. M. & Brammar, W. J. *Gene* **10**, 249-259 (1980).
- Yanisch-Perron, C., Vieira, J. & Messing, J. *Gene* **33**, 103-119 (1985).
- Buikema, W. J., Klingensmith, J. A., Gibbons, S. L. & Ausubel, F. M. *J. Bact.* **169**, 1120-1126 (1987).
- Noti, J. D., Folkerts, O., Turken, A. N. & Szalay, A. A. *J. Bact.* **167**, 774-783 (1986).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Vogelstein, B. & Gillespie, D. *Proc. natn. Acad. Sci. U.S.A.* **76**, 615-619 (1979).
- Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
- Tumer, N. E., Robinson, S. J. & Haselkorn, R. *Nature* **306**, 337-342 (1983).

T-DNA integration and expression in a monocot crop plant after induction of *Agrobacterium*

Willi Schäfer*, Andrea Görz & Günter Kahl†

Department of Biology, Plant Molecular Biology Group, University of Frankfurt, Siesmayerstraße 70, 6000 Frankfurt am Main, FRG

Agrobacterium tumefaciens induces tumorous proliferations on dicotyledonous plants by transferring a specific region (T-DNA) of its tumour-inducing (Ti) plasmid into wound-activated plant cells, where it becomes integrated into nuclear DNA¹ and expressed². The *vir* region of the Ti plasmid is essential for T-DNA transfer^{3,4} and is induced by specific wound substances from plants⁵⁻⁹. *Agrobacterium* is therefore a powerful tool for gene transfer to plants, but with some exceptions¹⁰⁻¹⁴ this method does not seem to be applicable to the commercially important monocotyledonous crops. We report here transformation of the monocotyledonous crop plant *Dioscorea bulbifera* (yam) with *agrobacteria* preincubated by wound substances from dicotyledons, leading to crown gall tumour formation. Moreover, we present direct evidence for the integration of T-DNA into the nuclear genome of this monocotyledonous crop.

* Present address: Department of Plant Pathology, Cornell University, Ithaca, New York, USA.

† To whom correspondence should be addressed.

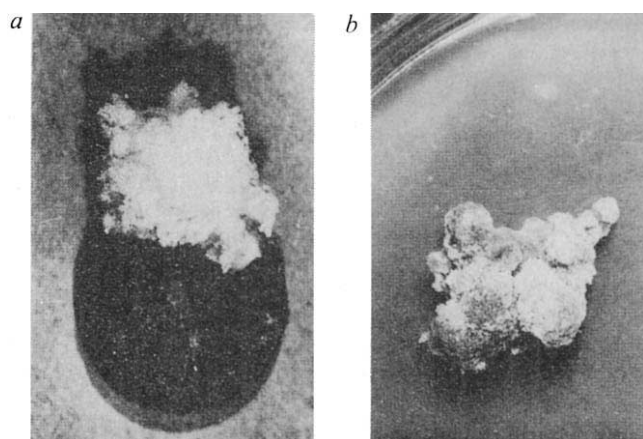


Fig. 1 Formation of crown gall tumour on a disk of *Dioscorea bulbifera* (Yam) bulbil tissue, 90 days after infection with induced *Agrobacterium tumefaciens*, strain C58 (a) and sterile tumour tissue culture on hormone-free medium 10 months after transfer (b). Tumours were ~1 cm (a) and 3 cm (b) in diameter.

Methods. For induction, *Solanum tuberosum* tuber tissue was cut into small pieces of about 2×2×2 mm, thoroughly washed with sterile tap water and 50 g batches of it incubated in 30 ml MS medium (Murashige and Skoog salts²¹ and 3% sucrose supplemented with 0.018% K₂HPO₄, 0.01% inositol, 0.0001% biotin, pH 5.5) under constant shaking at 25 °C for 1.5 h. The incubation mixture was filtered through several layers of Miracloth and the filtrate added to late log-phase *A. tumefaciens* C58 cells to a final concentration of 10¹¹ bacteria ml⁻¹. Yam bulbils were peeled, surface-sterilized in 0.5% NaOCl for 20 min, and cylinders punched out of the central part with a cork borer. Disks 2 mm thick and 1.5 cm in diameter were prepared from the cylinders and placed onto 1% MS agar in 150-mm Petri dishes, where they were inoculated with 10⁹ bacteria in 10 µl medium. The Petri dishes were incubated at 28 °C in the dark. Small white protuberances appeared on inoculated disks, but not on control disks (those without bacteria or treated with uninduced *A. tumefaciens* C58 or with the cured strain C58C1) after about five weeks. Two to three months after inoculation tumours (in A) were removed from the yam disks, cut into pieces, washed in MS medium supplemented with 0.5 mg ml⁻¹ cefotaximum (Claforan, Hoechst) and placed onto 0.8% agar in MS, containing 0.5 mg ml⁻¹ Claforan and grown at 25 °C under 8 h dark/16 h light regime (1,500–2,000 lux). Only few of the tumours survived this treatment and those that did which grew very slowly, but were sterile, were taken for DNA extraction after 10 months of cultivation.

Infection of wounds on dicotyledonous plants with virulent *A. tumefaciens* leads to the transfer, integration and expression of T-DNA in the infected plant cells which then proliferate into crown gall tumours. This infection mechanism does not seem to work properly with monocotyledonous plants (monocots). The inability of *Agrobacterium* to induce tumours on monocots or, alternatively, the inability of monocots to respond to infection by forming tumours has not been explained satisfactorily¹⁵. It has been suggested that no T-DNA transfer occurs either because the bacteria cannot attach to the cell walls of monocots¹⁶ or because of an abnormal auxin–cytokinin balance in monocot cells¹⁷. It is also possible that the transformation frequency in monocots is so low that it does not result in symptom formation¹⁴. However, one case of tumour formation has been demonstrated with the monocot *Asparagus officinalis*¹¹ and expression of opine genes has been found in infected wound sites of *Chlorophytum capense* (Liliaceae), *Narcissus* cv Paperwhite (Amaryllidaceae)¹⁰ and *Zea mays* (Gramineae)¹². The presence of transforming DNA in the recipient genome has not been demonstrated, however.

As an alternative explanation for the failure of *Agrobacterium* to transform monocots we hypothesized that there might be

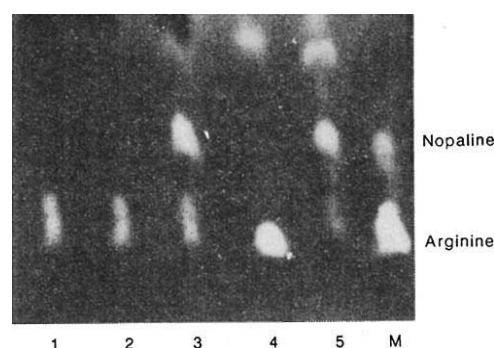


Fig. 2 Detection of nopaline in *D. bulbifera* tumours generated by induced *A. tumefaciens* C58 according to (ref. 18). Two months after inoculation with *A. tumefaciens* C58 the tumours (induced condition) or the surface of the *Dioscorea* bulbil disks (non-induced condition) were removed and 0.1 g extracted for nopaline, without preincubation in arginine. The compounds were separated using paper electrophoresis and stained with phenanthrene quinone reagent. M, nopaline and arginine markers. Lane 1, surface tissue of a *Dioscorea* bulbil disk treated with MS and exudate from wound-healing potato tuber tissue, but no agrobacteria. Lane 2, surface tissue of a bulbil disk inoculated with non-induced *A. tumefaciens* C58. Lane 3, crown gall tumours of *Dioscorea* bulbil disks inoculated with *A. tumefaciens* C58, induced with potato tuber wound exudates. Lane 4, surface tissue of a bulbil disk inoculated with induced *A. tumefaciens*, strain C58 C1. Lane 5, crown-gall tumours of *Solanum tuberosum* tuber disks inoculated with non-induced *A. tumefaciens* strain C58.

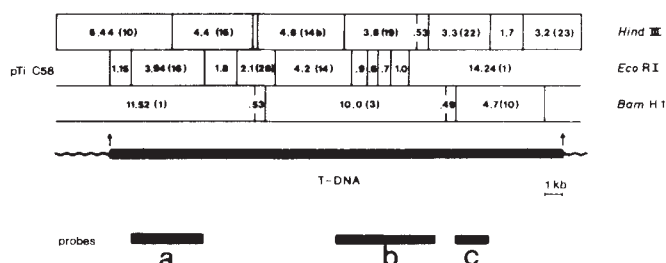


Fig. 3 Restriction map of the T-region of pTiC58 (ref. 22), the extent of T-DNA and the probes used in Southern blotting experiments. The number of each fragment is given in parentheses and the size in kilobases (kb). Probes a, b and c were derived from pGV0342 (ref. 22), pGV0354 (ref. 22), and pG4s (ref. 23), respectively (from Professors M. van Montagu, Gent and Jeff Schell, Cologne).

insufficient amounts of, or no, specific wound substances produced in these plants which could induce the vir regulon^{8,9}, a set of *trans*-acting sequences of the Ti plasmid involved in T-DNA excision and transfer. Because the monocots are so important as crop plants and thus desirable targets for plant biotechnology, we started an intense research programme to transform *Dioscorea bulbifera* (yam), a monocot crop plant of considerable commercial interest. (Note the distinction from sweet potato (*Ipomoea*), also called yam in some countries.) Our strategy involved induction of bacteria with dicot wound substances before infection, by analogy with the work of Zambryski^{6,8} and Schilperoort and colleagues⁵.

D. bulbifera bulbils were obtained from the Senckenberg Botanical Garden, Frankfurt am Main. Bulbil disks were infected with *A. tumefaciens* strain C58 wt, which had been treated with wound exudates of the dicot *Solanum tuberosum* L. (potato). Infection resulted in swellings of the disk surface after 4 weeks, in small white protuberances a week later, and typical crown gall tumours after about 10 weeks. At this time 35% of the inoculated disks showed at least one tumour, and 50% more than one. The *Dioscorea* tumours (Fig. 1a) proliferate more slowly than potato tumours, and the same is true for a tumour

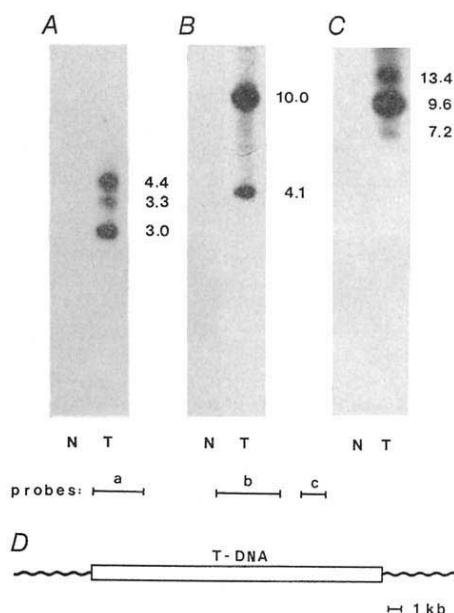


Fig. 4 Southern blot hybridization analysis of T-DNA in *D. bulbifera* tumours (A–C). Genomic DNA isolated from nontransformed *Dioscorea bulbifera* bulbil tissue (N) and sterile bulbil tumour culture (T) was cut and probed as follows: A, *Hind*III and probe a (left border); B, *Bam*HI and probe b (internal fragment); C, *Eco*RI and probe c (right border); D, diagram of probes used. **Methods.** DNA was isolated from nuclei using a modified Willmitzer–Wagner method²⁴. Tumour tissue was ground under liquid N₂ using mortar and pestle with homogenizing buffer (250 mM sucrose, 10 mM NaCl, 10 mM MES, 5 mM EDTA, 2.5 mM spermidine, 0.38 mM spermine, 20 mM β -MCE, 0.6% NP 40, 0.5% BSA, 0.2 mM phenylmethylsulphonylfluoride (PMSF), pH 6.0) in a 1:10 ratio (w/v). The homogenate was filtered through two layers of Miracloth, centrifuged for 5 min at 2,000g and 4 °C and the pellet washed with 0.5 vol homogenizing buffer. The crude nuclei were layered onto 6 g of 1.25 M sucrose, 50 mM NaCl, 50 mM MES, 25 mM EDTA, 12.5 mM spermidine, 1.9 mM spermine, pH 6.0, added to 45 g Percoll and centrifuged at 4,000g for 10 min at 4 °C. The floating nuclei were suspended in 30 ml 10 mM EDTA, 30 mM Tris, pH 8.0 and again centrifuged as before. The pellet was resuspended in 16 ml 10 mM EDTA, 30 mM Tris, supplemented with proteinase K and sarkosyl to a final concentration of 0.5 mg ml⁻¹ and 1%, respectively. The suspension was heated to 60 °C for 5 min, incubated at 37 °C for 3 h, and after addition of 28 g CsCl and 400 μ l ethidium bromide (10 mg ml⁻¹) centrifuged in a VTi 50 rotor (16 h, 45,000 r.p.m., 15 °C). The DNA was recentrifuged, ethidium bromide extracted with water-saturated *n*-butanol and the DNA dialysed against 5 mM Tris, 0.25 mM EDTA, pH 8.0. The resulting DNA of more than 50 kb was used for digestion and electrophoresis. Cut DNA was separated on 0.8% (w/v) agarose gels and 30 μ g of total plant DNA were hybridized to [α -³²P]dCTP-labelled probes with specific activities 1.0–3.0 $\times 10^9$ c.p.m. per μ g DNA, prepared by the oligo-labelling method²⁵. Sizes of hybridizing fragments are given in kb. Probes as in Fig. 3. Markers, *Hind*III or *Pst*I digests of wild-type λ DNA.

culture on solid media (Fig. 1b). Control disks inoculated with the induction medium from potato only showed no proliferation at all, nor did disks infected with uninduced bacteria, or with cured strain C58 C1 when induced, form any tumours.

As T-DNA genes encode enzymes for the production of opines which are not normally synthesized by the host organism, the occurrence of such opines in plant cells is evidence for the existence and expression of T-DNA. We collected tumours of *D. bulbifera* and assayed for nopaline¹⁸. Nopaline was found in all tumours in vast amounts (Fig. 2, lane 3). Wounded bulbil tissue, bulbil tissue treated with non-induced *A. tumefaciens* or bulbil tissue treated with cured strain C58 C1 did not produce any nopaline (Fig. 2, lanes 1, 2 and 4).

We conclude that once transferred, the T-DNA is expressed in *Dioscorea* just as it is in dicots, resulting in the production of nopaline synthase and the formation of crown gall tumours. In nature the initial step of *vir* induction seems to be prevented by the inability of monocotyledons to synthesize inducing wound substances. This insufficiency may be overcome with wound exudates from dicotyledonous donors. We expect that the host range of *A. tumefaciens* will be considerably expanded through the use of bacteria with induced *vir* genes.

The occurrence of nopaline in tumours is evidence for the transfer of the T-DNA and the expression of its *nos* gene, but not of integration¹⁹. We therefore investigated the integration of T-DNA into the nuclear DNA of the *Dioscorea* host cells using Southern analysis (Fig. 4; probes used are explained in Fig. 3). *Dioscorea* DNA could hardly be cut with restriction enzymes if extracted and purified according to standard methods, so we had to devise new protocols for its isolation (Fig. 4 legend). Total DNA was cut with either *Hind*III, *Bam*HI or *Eco*RI and hybridized to probe a (left border), probe b (internal sequences) and probe c (right border), respectively (Fig. 4). It is clear that T-DNA is integrated into the nuclear DNA of this monocotyledonous plant. When the DNA was cut with *Hind*III and probed with probe a (Fig. 4), three T-DNA homologous signals appear, migrating at 3.0, 3.3 and 4.4 kilobases (kb). The latter signal represents an internal fragment (Fig. 3). The two additional signals indicate that two T-DNA copies have been integrated at different sites in the nuclear DNA. *Bam*HI-restricted DNA probed with probe b was expected to give a 10-kb T-DNA signal only. Figure 4C shows a second signal at 4.1 kb. We interpret this band as indication that we are dealing with T-DNA copies of different internal structure. Hybridization of *Eco*RI-cut DNA probed with probe c gives three signals at 7.2, 9.6 and 13.4 kb indicating three different integration sites in the *Dioscorea* genome. We conclude that two full-length copies are to be found together with a truncated T-DNA copy consisting of the right border and an internal fragment.

These results show that infection of wound sites on the monocotyledon *D. bulbifera* with induced *A. tumefaciens* strain C58 leads to tumour formation. For successful tumorigenesis induction of the *vir* regulon seems to be one of the prerequisites, which in our experiments is apparently brought about by treatment of bacteria with wound exudates from dicotyledonous plants (potato). Moreover, at least some of the T-DNA genes are expressed, because transformed cells proliferate and nopaline accumulates. We present here direct evidence that T-DNA is integrated into the nuclear DNA of a monocotyledon, in this case *Dioscorea*, where it occurs as two full-length copies with different integration sites and an additional, truncated T-DNA copy.

Our results suggest that *A. tumefaciens* can be used for the genetic engineering of monocots as well as dicots, including such important crop plants as *D. bulbifera*²⁰. High transformation efficiency may be obtained by induction with wound substances from dicotyledons.

We thank Jeff Schell and Marc Van Montagu for the T-DNA clones, the Fritz Thyssen Foundation for financial support and Mrs S. Kost for help with the manuscript.

Received 6 April; accepted 21 April 1987.

- Kahl, G. & Schell, J. *Molecular Biology of Plant Tumors* (Academic, New York, 1982).
- Willmitzer, L. *et al.* *Cell* **32**, 1045–1056 (1983).
- Yanofsky, M. F. *et al.* *Cell* **47**, 471–477 (1986).
- Stachel, S. E. & Zambryski, P. C. *Cell* **47**, 155–157 (1986).
- Okker, R. J. H. *et al.* *Nature* **312**, 564–566 (1984).
- Stachel, S. E., Messens, E., van Montagu, M. & Zambryski, P. *Nature* **318**, 624–629 (1985).
- Stachel, S. E. & Nester, E. W. *EMBO J.* **5**, 1445–1454 (1986).
- Stachel, S. E., Nester, E. W. & Zambryski, P. *Proc. natn. Acad. Sci. U.S.A.* **83**, 379–383 (1986).
- Stachel, S. E., Timmerman, B. & Zambryski, P. *Nature* **322**, 706–712 (1986).
- Hooikaas-van Sloten, G. M. S., Hooikaas, P. J. J. & Schilperoord, R. A. *Nature* **311**, 763–764 (1984).

11. Hernsteens, J. P., Thia-Toong, L., Schell, J. & van Montague, M. *EMBO J.* **3**, 3039-3041 (1984).
12. Graves, A. C. F. & Goldman, S. L. *Pl. molec. Biol.* **7**, 43-50 (1986).
13. Grimsley, N., Hohn, B., Hohn, Th. & Walden, R. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3282-3286 (1986).
14. Grimsley, N., Hohn, Th., Davies, J. W. & Hohn, B. *Nature* **325**, 177-179 (1987).
15. DeCleene, M. & DeLey, J. *Bot. Rev.* **42**, 389-466 (1976).
16. Rao, S. S., Lippincott, B. B. & Lippincott, J. A. *Physiol. Plant.* **56**, 374-380 (1982).
17. Krens, F. A., Willems, G. J. & Schilperoort, R. A. in *Structure and Function of Plant Genomes* (eds Ciferri, O. & Dure, L.) 387-408 (Plenum, New York, 1983).
18. Otten, L. A. B. M. & Schilperoort, R. A. *Biochim. biophys. Acta* **527**, 497-500 (1978).
19. Christou, P., Platt, S. G. & Ackerman, M. C. *Pl. Physiol.* **82**, 218-221 (1986).
20. Ammirato, P. V. in *Handbook of Plant Cell Culture* Vol. 3 (eds Ammirato, P. V., Evans, R., Sharp, Y. & Yamada, Y.) 327-354 (Macmillan, London, 1983).
21. Murashige, T. & Skoog, F. *Physiologia Plant.* **15**, 473-497 (1962).
22. Depicker, A. et al. *Plasmid* **3**, 193-211 (1980).
23. Inzé, D. et al. *Molec. gen. Genet.* **194**, 265-274 (1984).
24. Willmitzer, L. & Wagner, K. *Expl. Cell Res.* **135**, 69-77 (1981).
25. Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6-13 (1983).

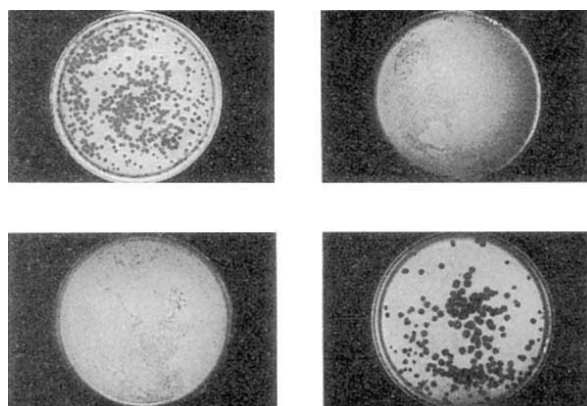


Fig. 1 Transfection of *M. smegmatis* spheroplasts with mycobacteriophage D29 DNA. Top left, *M. smegmatis* spheroplasts transfected with D29 DNA; top right, same but D29 DNA treated with DNase; bottom left, same as top left but plated on medium lacking sucrose. Bottom right, spheroplasts from D29-resistant *M. smegmatis* transfected with D29 DNA.

Methods. Spheroplasts were prepared as described below from the *M. smegmatis* strain mc²6, a single colony isolate which is the predominant colony type isolated from the ATCC 607 *M. smegmatis* stock culture and which forms orange rough colonies on regeneration media²⁰. A second strain, mc²11, was isolated as a spontaneous D29-resistant isolate of the ATCC 607 *M. smegmatis* stock culture when 10⁸ cells were mixed with 3 × 10⁸ PFU of D29 and plated on tryptic soy agar plates. D29-resistant colonies arose at a frequency of 10⁻⁷. Spheroplasts of mc²6 were mixed with 1 µg of D29 DNA, and subsequently one tenth of that mixture was plated on tryptic soy agar plates with or without 0.5 M sucrose and then overlaid with the appropriate soft agar containing 10⁸ mc²6 cells. The DNase treatment was performed by adding DNase I (sigma) at final concentration of 50 µg ml⁻¹ to the D29 DNA. Equivalent amounts of mc²11 spheroplasts were used in the experiment above, but then subsequently overlaid with mc²6 cells to assay plaque forming units. Preparation of mycobacteriophage and phage DNA: Plate lysates of D29 (ref. 21) were prepared using mc²6 cells and tryptic soy agar as described previously²². Phage were purified on two CsCl equilibrium gradients followed by extensive dialysis against MP buffer (10 mM Tris, pH 7.6, 100 mM NaCl, 10 mM MgSO₄, 2 mM CaCl₂). DNA was purified from phage by proteinase-K treatment, phenol-chloroform extraction and extensive dialysis against TE buffer. Spheroplasts of *M. smegmatis* were prepared as for *Streptomyces*²⁰ using media for the preparation of lysozyme-sensitive cells as for *M. smegmatis*²³. Briefly, 40 ml of mc²6 cells were grown in tryptic soy broth containing 0.2% Tween-80 at 37 °C to an A₆₀₀ of 0.2, at which time glycine was added to a final concentration of 1%. The cells were incubated for an additional 16 h, harvested by centrifugation, washed twice with 10.3% sucrose, and resuspended in P (protoplast) buffer²⁰ containing 2 mg ml⁻¹ lysozyme. After a 2-h incubation at 37 °C, the spheroplasts were pelleted and resuspended in 10 ml P buffer. For each transfection, 2.5 ml of the spheroplast suspension was pelleted, the supernatant fluid was carefully decanted and the spheroplasts were resuspended in the remaining drop of P buffer. Following the addition of 1 µg of DNA, 0.5 ml of 25% polyethylene glycol (PEG-1000) prepared in P buffer was mixed with the spheroplasts and DNA. This mixture was diluted with 5 ml P buffer, and the spheroplasts were pelleted and resuspended in 1 ml of P buffer. Samples were transferred to tryptic soy agar with 0.5 M sucrose. The plates were then overlaid with 3.0 ml of soft tryptic soy-sucrose agar and incubated at 37 °C. The plaques were counted after 24 h incubation.

cies of 10³-10⁴ plaque-forming units (PFU) per µg D29 DNA were routinely obtained. These plaques were shown to be the result of transfection of *M. smegmatis* spheroplasts by three criteria: first, transfection was abolished by DNase; second, osmotic shock of treated cells prevented productive transfection and third, spheroplasts derived from a D29 phage-resistant mutant of *M. smegmatis* were transfected at frequencies comparable to those of the parent strain.

Introduction of foreign DNA into mycobacteria using a shuttle phasmid

William R. Jacobs Jr, Margareta Tuckman & Barry R. Bloom

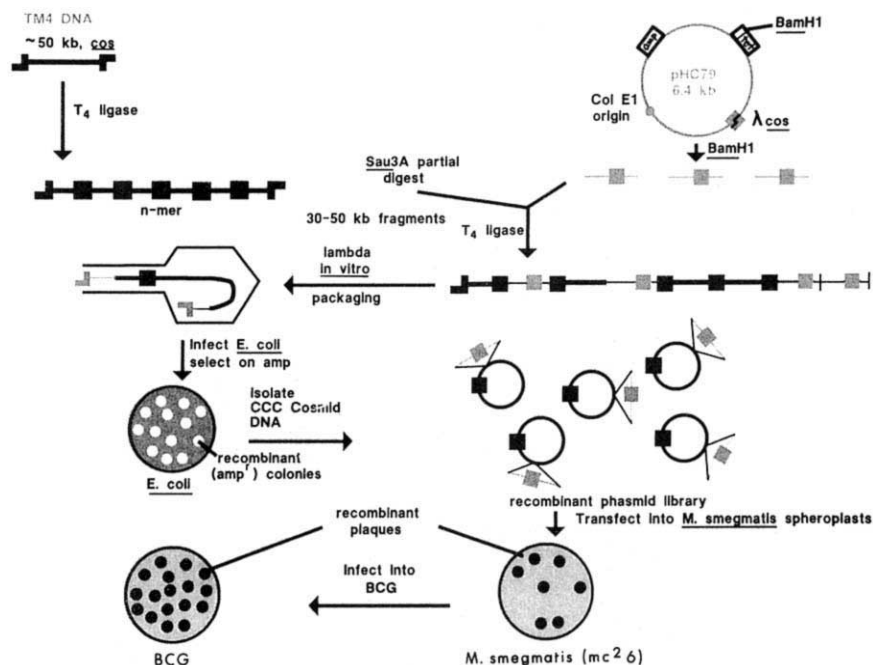
Department of Microbiology and Immunology, Albert Einstein College of Medicine, Bronx, New York 10461, USA

Mycobacteria are major pathogens of man and animals. There are ~10 million cases of tuberculosis world wide with an annual mortality of three million¹ people. Leprosy, caused by *Mycobacterium leprae*, afflicts over ten million people, primarily in developing countries². *M. tuberculosis* and mycobacteria of the *M. avium-intracellulare-scrofulaceum* (MAIS) group are major opportunistic pathogens of patients with acquired immune deficiency syndrome (AIDS)³. *M. paratuberculosis* is the cause of Jöhne's disease in cattle. Yet, BCG (Bacille Calmette-Guerin), an avirulent strain of *M. bovis*, is the most widely used human vaccine in the world, having been administered to about 2.5 × 10⁹ people since 1948 (ref. 4). BCG was highly protective against tuberculosis in England⁵, but has been found not to be effective in preventing pulmonary tuberculosis in adults in Southern India⁶. We have initiated studies to develop the methodology for efficient gene transfer in mycobacteria. We have constructed recombinant shuttle phasmids which are chimaeras containing mycobacteriophage DNA into which an *E. coli* cosmid is inserted. They can replicate in *E. coli* as plasmids and in mycobacteria as phages, and transfer DNA across both genera. These shuttle vectors permit for the first time the introduction of foreign DNA by infection into *M. smegmatis* and BCG. By introducing and ultimately expressing genes for protective antigens for a variety of pathogens, it may be possible to develop cultivatable mycobacteria into useful multivaccine vehicles.

To develop a system for the manipulation of DNA in mycobacteria it was first necessary to devise an efficient means of transferring DNA into the bacillus. We adapted the technology for preparing protoplasts developed for the related actinomycete, *Streptomyces*^{7,8} for use with *M. smegmatis*, with the addition of polyethylene glycol to facilitate entry of DNA molecules into bacterial protoplasts⁹. The system chosen to evaluate optimum conditions for DNA transfer into mycobacteria was the transfection of DNA from lytic mycobacteriophages, because the results obtained were quantitative and readily apparent within 24 h. Transfection experiments were initiated with DNA from mycobacteriophage D29, which propagates on a wide variety of mycobacteria and forms large clear plaques on *M. smegmatis*. Restriction analyses of ligated and unligated D29 DNA demonstrated that the phage genomic DNA was double stranded, ~50 kilobases (kb) in size, and possessed cohesive ends (data not shown).

The results of transfection of *M. smegmatis* spheroplasts by mycobacteriophage D29 DNA are illustrated in Fig. 1. Efficient-

Fig. 2 Diagram of the construction of the shuttle phasmids. TM4 phage DNA was ligated at a concentration of $250 \mu\text{g ml}^{-1}$ and then partially digested with *Sau3A* to obtain fragments that averaged 30–50 kb. These fragments were ligated at a 1:2 molar ratio of TM4 fragments to pHCT9 that had been cleaved with *Bam*HI and alkaline phosphatased. The packaging of an aliquot of this ligation with *in vitro* packaging mix (Gigapack Plus, Stratagene) and subsequent transduction into ER1381 (*hsdR mcrA⁺ mcrB⁺*, E. Raleigh, personal communication), yielded 10^6 ampicillin colonies per μg of TM4 DNA insert, when plated on L-agar containing ampicillin at $50 \mu\text{g ml}^{-1}$. A pool of 40,000 ampicillin-resistant clones was prepared by homogenizing colonies in L-broth from the selection plates with a glass spreader. Plasmid was isolated from the pool of clones by alkaline-SDS treatment²⁴ followed by phenol-chloroform extraction and concentration with ethanol. Covalently closed circular plasmid DNA was transfected into *mc*²6 spheroplasts as in Fig. 1. The plaques were screened for the presence of pHCT9 by performing plaque lifts²⁵ and Biotrans nylon membranes (ICN Radiochemicals). The membranes were hybridized with pHCT9 DNA nick-translated with ³²P-dCTP, and autoradiography was carried out.



Our immediate aim was to develop a vector that would permit both the manipulation and amplification of mycobacterial DNA constructs in *E. coli*, and subsequent transfer and replication in a variety of mycobacteria. In particular, we wanted to be able to introduce DNA into fast-growing non-pathogens such as *M. smegmatis*, but also into slow-growing mycobacteria, such as BCG and *M. tuberculosis*. Although plasmids have been found in some mycobacterial strains in the MAIS complex and *M. fortuitum*^{10–13}, none have yet been described which replicate in *M. smegmatis*, BCG, or *M. tuberculosis* and, with one exception, none possess selectable markers. In contrast, a variety of phages that replicate in *M. smegmatis*, BCG, and *M. tuberculosis* have been described and long used for typing isolates. We reasoned that the best strategy might be to construct a vector that would replicate as a plasmid in *E. coli* and as a phage in mycobacteria. A λ -ColE1 vector with dual properties in *E. coli* was previously designated a phasmid¹⁴. Our aim was to develop bifunctional vectors which were similar to those described for *Streptomyces*¹⁵, except that they could also be packaged *in vitro* into λ phage, which would permit highly efficient transfer of DNA into *E. coli* and subsequently into mycobacteria. For this purpose we selected the mycobacteriophage TM4, a temperate phage isolated from *M. avium*¹⁶, which we had found replicates in *M. smegmatis*, BCG, and *M. tuberculosis*. This phage has a double-stranded DNA genome of ~50 kilobases (kb) and possesses cohesive ends (Fig. 3).

Our strategy of using cosmid cloning¹⁷ to introduce an *E. coli* plasmid replicon into a non-essential region of phage TM4 is illustrated in Fig. 2. The cohesive ends of TM4 genomes were ligated together to form long concatamers which were then partially digested with *Sau3A* to generate a set of DNA fragments that contained the entire TM4 genome or genomes with small deletions, but had been cleaved at any of the *Sau3A* sites in the genome. These DNA fragments were then ligated to the cosmid pHCT9 (ref. 18) which had been cleaved with *Bam*HI. To select for recombinant molecules of the appropriate size, the ligation mixture was packaged into phage λ heads *in vitro*. The resulting phage particles were transduced into *E. coli* and clones containing recombinant pHCT9::TM4 DNA molecules were selected on media containing ampicillin. Covalently closed circular plasmid DNA was isolated from 40,000 pooled ampicillin-resistant colonies. This library, containing recombinant

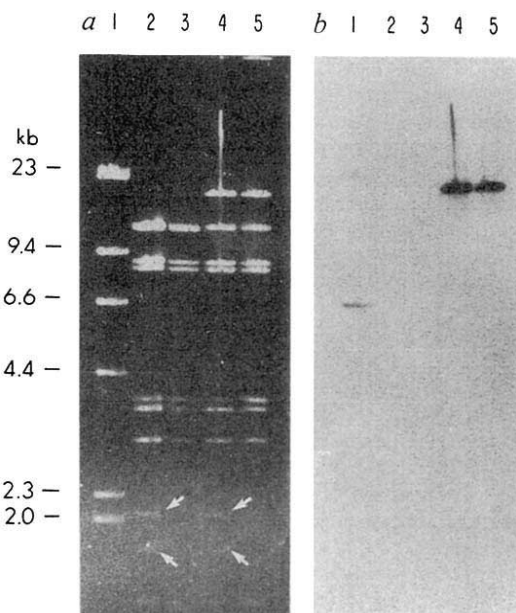
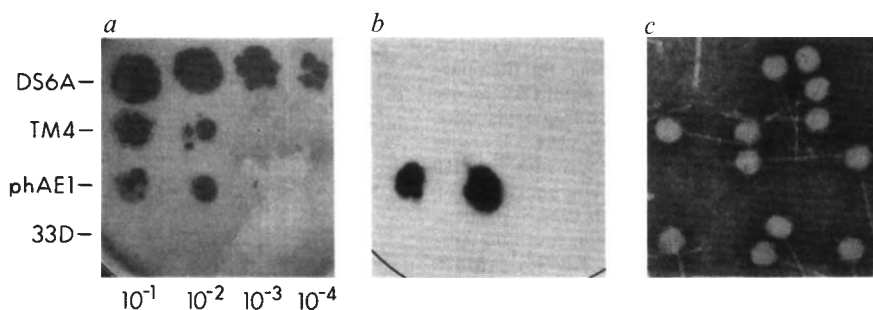


Fig. 3 *a*, Ethidium bromide stained 0.7% agarose gel of TM4 and phAE1 DNAs digested with *Kpn*I. Lane 1, λ DNA digested with *Hind*III. Lanes 2 and 3, TM4 DNA that was unligated (lane 2) or ligated (lane 3) before digestion, cut with *Kpn*I. Lanes 4 and 5, phAE1 DNA that was isolated from phage particles propagated on *M. smegmatis* (lane 4) and phAE1 that had been isolated from *E. coli* cells as a plasmid (lane 5). Arrows, unannealed 2.1- and 1.8-kb TM4 cohesive end fragments, present in DNA preparations from mycobacteriophage. These bands are absent when phage DNAs are ligated before digestion or when shuttle phasmid DNA is isolated as a plasmid. *b*, Southern blot analysis of phAE1 using pHCT9 as a probe. An autoradiograph of *a*, after performing a Southern blot²⁶ onto nylon membrane and probing with pHCT9 DNA that had been nick-translated with ³²P-dCTP.

molecules of TM4 genomes into which pHCT9 cosmid DNA had been randomly inserted in *Sau3A* sites around the TM4 genome, was transfected into *M. smegmatis* spheroplasts to select for TM4 phages which had pHCT9 inserted in non-essential regions. The transfection yielded $100 \text{ PFU } \mu\text{g}^{-1}$ of plasmid DNA. Among 4,000 plaques screened for the presence of

Fig. 4 Replication of phAE1 on BCG and ultrastructural analysis of phAE1 mycobacteriophage particles. *a*, Comparison of lysis of the Glaxo vaccine strain of BCG by DS6A (ref. 22), a mycobacteriophage known to form plaques on *M. tuberculosis* and BCG, but not other mycobacteria; phage 33D, known to form plaques on *M. tuberculosis* and not BCG²⁷; and phages TM4 and phAE1. Titres of phage (PFU ml⁻¹) used at 10⁻¹ dilution were: DS6A, 2 × 10⁶ on *M. tuberculosis* strain H37Ra; 33D, 2 × 10⁶ on *M. smegmatis* strain mc²6; TM4, 3 × 10⁸ on mc²6; and phAE1, 3 × 10⁸ on mc²6. Dilutions of phages (5 µl) were spotted on a soft agar overlay containing 10⁸ BCG cells prepared as described by Jones *et al.*²⁸ and the resulting lysis was photographed after incubation for 10 d at 37 °C. Similar results were obtained for the Danish and Pasteur strains of BCG. Both TM4 and phAE1 form plaques 5–6 logs of ten less efficiently on BCG than on mc²6. This decrease is most probably a result of restriction of phage grown on *M. smegmatis* by BCG, because phage prepared from BCG phage lysates give equal efficiencies when plaqued on BCG and mc²6. *b*, Presence of pHc79 in phAE1 replicated on BCG. A plaque lift was performed on the plate in *a* as described above, hybridized with ³²P-labelled pHc79 DNA, and the resulting autoradiograph is shown. *c*, Electron micrograph of shuttle phasmid phAE1 mycobacteriophage particles. CsCl-purified phage particles were placed on carbon coated, Parlodion-coated grids, blotted and washed with one drop of 1% phosphotungstic acid. Electron micrographs were taken using a JEOL 1200EX electron microscope at 80 kV, magnification ×30,000.



pHC79, only 10 were positive. Following plaque purification and propagation on *M. smegmatis* cells, one such phage (phasmid pHAE1) was studied in more detail. Restriction fragment analysis of DNAs isolated from a number of independent phages that had resulted from the transfection of the pHc79::TM4 DNA library into *M. smegmatis*, and that did not hybridize with pHc79, revealed them to be identical to TM4 DNA. We believe these phages resulted from a recombination event occurring in transfected cells containing two or more pHc79::TM4 molecules, yielding a wild-type TM4 genome.

DNA isolated from phAE1 phage particles grown on *M. smegmatis* were subsequently ligated at the TM4 cohesive ends and packaged *in vitro* into bacteriophage λ heads using the λ cohesive ends, yielding particles which efficiently transduced ampicillin resistance to *E. coli* cells. The ability of these shuttle phasmids to use the λ *in vitro* packaging system will greatly facilitate subsequent cloning of genes into these vectors. Gel analysis of TM4 and phAE1 DNAs digested with *Kpn*I revealed the presence of a new 18-kb fragment in phAE1 with a subsequent loss of one of the 12-kb doublets in TM4 (Fig. 3a). Southern analysis confirms that the cosmid pHc79 is present in the 18-kb fragment (Fig. 3b), indicating that pHc79 has been inserted in a non-essential region of a 12-kb *Kpn*I fragment of the TM4 genome. The phAE1 DNAs, isolated from mycobacteriophage particles or plasmid DNA from *E. coli*, showed identical restriction patterns, except for the presence of the unannealed TM4 cohesive ends in the phage DNA (Fig. 3a). The phAE1 mycobacteriophage particles have hexagonal heads that average 50 µm in diameter and long tails of 180–220 µm in length with a disk-like baseplate present on many of the tails (Fig. 4c). This structure is very similar to the parent TM4 phage¹⁴.

One of the goals of this research is to enable us to introduce DNA encoding important antigens into BCG vaccine strains. However, transformation of slow-growing mycobacterial strains, including BCG, has not yet been achieved. It is therefore gratifying that the shuttle phasmid phAE1 obtained from *M. smegmatis*, like its parent TM4, is able to infect, and replicate in, three different *M. bovis*-BCG vaccine strains tested, the Glaxo, Pasteur and Danish BCGs (Fig. 4a and b). Current efforts are directed towards developing conditions for stable expression of cloned genes in mycobacterial cells.

We believe that the ability to infect BCG vaccine strains with the shuttle phasmids provides a novel means of introducing cloned genes into slow-growing mycobacteria and offers the possibility of developing recombinant mycobacterial multivaccine vehicles. This could be achieved if protective antigens from a variety of pathogens, particularly those requiring T-cell

memory or effector function, could be introduced and stably expressed. BCG vaccine has been administered to billions of people over the past 35 years with extraordinarily few side-effects and a mortality rate of 60 per billion of those treated¹⁹. BCG is probably the most potent immunological adjuvant known, especially for engendering cell-mediated immunity. It is the only vaccine currently given at birth and a single inoculation induces cell-mediated immunity for 5–50 years, at a cost of only \$0.055 per inoculation. We hope as well that the methodology described here will facilitate the development of a molecular genetic system in pathogenic mycobacteria that will permit better understanding, and perhaps molecular modification, of their pathogenicity.

We thank Drs Richard Young and Ronald Davis for their encouragement, advice and criticism; Wilbur Jones, Jack Crawford, Don Auclair and Elizabeth Raleigh for various phage and bacterial strains and advice on using them; Jane Fant for the electron microscopy; Scott Snapper for assistance and valuable discussion; and Drs Roy Curtiss III and Josie Clark-Curtiss, the *Streptomyces* group of the John Innes Institute (W.R.J.), and Dr George Stark (B.R.B.) for providing us with sufficient backgrounds in molecular biology to undertake these experiments. This work was supported by the Immunology of Tuberculosis Committee of the WHO Programme on Vaccine Development and the Immunology of Leprosy Committee of the UNDP-WHO-World Bank Special Programme for Research and Training in Tropical Diseases, the Irvington House Institute for Medical Research. The Weinstock professorship (B.R.B.) and USPHS Grants from NIH.

Received 2 March; accepted 5 May 1987.

1. Joint International Union Against Tuberculosis and World Health Organization Study Group *Tubercle* **63**, 157–169 (1982).
2. Bloom, B. R. & Godal, T. *Rev. infect. Dis.* **5**, 765–780 (1983).
3. Centers for Disease Control *Morbidity and Mortality Weekly Report* **34**, 774 (1986).
4. United Nations Children's Emergency Fund (UNICEF) *The State of the World's Children* 131 (Oxford University Press, 1985).
5. Hart, P. D. & Sutherland, I. *Br. Med. J.* **2**, 293–295 (1977).
6. Tuberculosis Prevention Trial, Madras *Ind. J. med. Res.* **72** (suppl.), 1–74 (1980).
7. Okanishi, M., Suzuki, K. & Umezawa, H. *J. gen. Microbiol.* **80**, 389–400 (1974).
8. Hopwood, D. A. & Wright, H. M. *Molec. Gen. Genet.* **162**, 307–317 (1978).
9. Bibb, M. J., Ward, J. M. & Hopwood, D. A. *Nature* **274**, 398–400 (1978).
10. Crawford, J. T. & Bates, J. H. *Infect. Immunology* **24**, 979–981 (1979).
11. Mizuguchi, Y., Fukunaga, M. & Taniguchi, H. *J. Bact.* **146**, 656–659 (1981).
12. Meissner, P. S. & Falkinham, J. O. III *J. Bact.* **157**, 669–672 (1984).
13. Labidi, A., David, H. L. & Roulland-Dussoix, D. *Annls Microbiologie* **136B**, 209–215 (1985).
14. Brenner, S., Cesareni, G. & Karn, J. *Gene* **17**, 27–44 (1982).
15. Suarez, J. E. & Chater, K. F. *Nature* **286**, 527–529 (1980).
16. Timme, T. L. & Brennan, P. J. *J. gen. Microbiol.* **130**, 2059–2066 (1984).
17. Collins, J. & Hohn, B. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4242–4246 (1978).
18. Hohn, B. & Collins, J. *Gene* **11**, 291–298 (1980).
19. Lotte, A. *et al. Adv. Tuberc. Res.* **32**, 107–193 (1984).
20. Hopwood, D. A. *et al. In Genetic Manipulation of Streptomyces-A Laboratory Manual* (John Innes Foundation, Norwich, 1985).

21. Froman, S., Will, D. W. & Bogen, E. *Am. J. Public Health* **44**, 1326-1334 (1954).
22. Redmond, W. B. & Cater, J. C. *Am. Rev. resp. Dis.* **82**, 781-787 (1960).
23. Udou, T., Ogawa, M. & Mizuguchi, Y. *J. Bact.* **151**, 1035-1039 (1982).
24. Birnboim, H. C. & Doly, J. *Nucleic Acids Res.* **7**, 1513-1525 (1979).
25. Benton, W. D. & Davis, R. W. *Science* **196**, 180-182 (1977).
26. Southern, E. *J. molec. Biol.* **98**, 503-517 (1975).
27. Jones, W. D. Jr *Tubercle, Lond.* **60**, 55-58 (1979).
28. Jones, W. D. Jr & Greenberg, J. *J. clin. Microbiol.* **7**, 467-469 (1978).

Optically active benzo[*c*]phenanthrene diol epoxides bind extensively to adenine in DNA

A. Dipple*, M. A. Pigott*, S. K. Agarwal†, H. Yagi†, J. M. Sayer† & D. M. Jerina†

* BRI-Basic Research Program, NCI-Frederick Cancer Research Facility, Frederick, Maryland 21701, USA

† Laboratory of Bioorganic Chemistry, National Institute of Diabetes and Digestive and Kidney Diseases, The National Institutes of Health Bethesda, Maryland 20892, USA

Reactions of diol epoxide metabolites of carcinogenic polycyclic aromatic hydrocarbons with DNA are thought to initiate the carcinogenic process^{1,2}. Although formation of a benzo[*a*]pyrene (BaP) diol epoxide-deoxyguanosine adduct has been held responsible for biological activity³⁻⁵, the more potent carcinogen, 7,12-dimethylbenz[*a*]anthracene (DMBA) binds extensively to deoxyadenosine residues in DNA, suggesting that hydrocarbon carcinogen-deoxyadenosine adducts may be instrumental in tumour initiation⁶⁻⁸. Because the bay region diol epoxides⁹ of benzo[*c*]phenanthrene (BcPh) are very active tumour initiators¹⁰, and the relative activities of the four configurationally isomeric 3,4-diol 1,2-epoxides (Fig. 1) are known¹¹, we examined their reactions with DNA. Each BcPh diol epoxide isomer exhibits a remarkable preference for covalent binding to DNA over hydrolysis, each yields a unique distribution of products with the nucleosides of DNA and each reacts extensively with deoxyadenosine residues in DNA. The relative tumour initiating activities of these stereoisomers is best reflected by the relative yields of one of the deoxyadenosine adducts formed.

Tetraol formation in the presence of DNA indicated that 60-75% of the four BcPh diol epoxides bound covalently to DNA¹² in contrast to just 6% for a BaP diol epoxide^{12,13}. DNA adduct formation has now been measured directly for one of the BcPh diol epoxide isomers. When calf thymus DNA (0.8 mg ml⁻¹) was treated with (-)-diol epoxide-2 (Fig. 1) at 0.097, 0.020 and 0.004 mg ml⁻¹, 61%, 67% and 67% respectively of the diol epoxide was recovered as nucleoside adducts, in agreement with estimates from tetraol formation. This remarkable reactivity towards DNA could contribute to the high tumorigenic activity of some of the BcPh diol epoxides but cannot be the predominant factor, because all four stereoisomers react extensively with DNA whereas their tumour-initiating activities range from low to high¹¹.

Chromatographic analyses indicated that at least four different diol epoxide-nucleoside adducts were formed from each isomer. The nucleoside of origin of the principal adducts (Fig. 1) was established by comparing their spectral and chromatographic properties with those of adducts formed in separate reactions of each diol epoxide with individual nucleotides. The UV spectra of the deoxyadenosine and deoxyguanosine adducts were particularly useful for initial structural assignments. Only 8 of the 16 spectra (Fig. 2) are shown (solid lines for adducts with a 1 subscript in Fig. 1 and broken lines for adducts with a 2 subscript) because the spectra of the four adducts derived from the (+)-enantiomer of each diol epoxide (dGuo₁, dGuo₂, dAdo₁ and dAdo₂, see Fig. 1 for nomenclature)

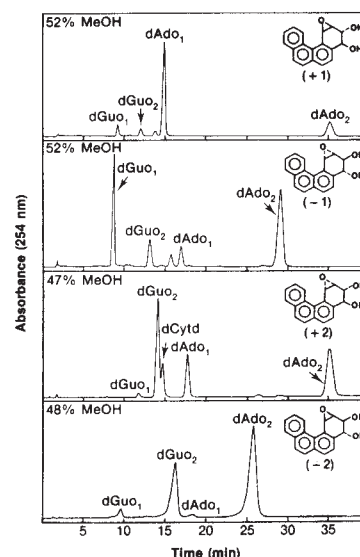


Fig. 1 Chromatographic separations of adducts formed in the reactions of configurationally isomeric BcPh 3,4-diol-1,2-epoxides with DNA. For diol epoxide-1, the benzylic hydroxyl group and the epoxide oxygen are *cis* and for diol epoxide-2, these groups are *trans*. To aliquots (0.5 ml) of a calf thymus DNA (Sigma) solution (0.8 mg ml⁻¹ in 0.05 M Tris-HCl buffer, pH 7) was added 0.05 ml of an acetone solution of the appropriate isomeric diol epoxide (2 mg ml⁻¹). After 5 h at 37 °C, each reaction mixture was diluted to 1 ml with 0.05 M Tris-HCl and was then extracted three times with 2 vol ethyl acetate and once with ether. Residual ether was removed from the aqueous solution in a stream of N₂. To convert the DNA to deoxyribonucleoside, each sample was incubated with deoxyribonuclease, snake venom phosphodiesterase and alkaline phosphatase⁶. The digest was then loaded on a Waters Sep-pak C₁₈ cartridge which was washed with water (20 ml) and 25% methanol (20 ml) before elution of the benzo[*c*]phenanthrene diol epoxide/deoxyribonucleoside adducts in methanol (2.6 ml). The components of the methanol eluates were then eluted isocratically in the aqueous methanol solvents indicated in Fig. 1 from an Altex Ultrasphere ODS column⁶. The adducts are labelled as deoxyguanosine (dGuo), deoxyadenosine (dAdo), or deoxycytidine (dCyt) adducts by comparison of retention times and UV absorption spectra with those for products from reactions of individual nucleotides (0.5 ml at 4.0 mg ml⁻¹ in 0.01 M Tris-HCl buffer, pH 7) with diol epoxide solution (0.05 ml at 2 mg ml⁻¹ in acetone). These reaction mixtures were extracted with organic solvents, the pH was raised to 9.0 and alkaline phosphatase (25 units ml⁻¹) was added before incubation overnight. Products were recovered on Sep-paks, as for the DNA reactions. Under the conditions used here, retention times in minutes with percentage of total products given in parenthesis for products dGuo₁, dGuo₂, dAdo₁ and dAdo₂, respectively were: (+)-diol epoxide-1, 9.5 (5.5%), 13.0 (5%), 14.9 (67%) and 36.1 (21%); (-)-diol epoxide-1, 8.4 (28%), 13.1 (10.5%), 16.3 (7%) and 28.2 (49%); (+)-diol epoxide-2, 11.5 (1.5%), 13.5 (35%), 16.3 (16.5%) and 32.1 (34.5%); (-)-diol epoxide-2, 10.4 (3%), 17.4 (30%), 19.6 (1.5%) and 27.6 (64%). As described elsewhere¹⁴, for diol epoxide-1 adducts dGuo₁ and dAdo₁ and diol epoxide-2 adducts dGuo₂ and dAdo₂, the epoxide ring has opened to yield *trans* products whereas for diol epoxide-1 adducts dGuo₂ and dAdo₂ and diol epoxide-2 adducts dGuo₁ and dAdo₁, the epoxide ring has opened to yield *cis* products.

can be superimposed on the spectra of the same adducts derived from the (-)-enantiomer. The members of each pair of adducts with identical spectra presumably differ from one another only in that there is a mirror-image relationship between the chiral centres on the BcPh residues in each case. The spectra are useful also for distinguishing deoxyadenosine adducts from deoxyguanosine adducts and for distinguishing the two isomeric products formed with each purine nucleoside by a given diol epoxide (Fig. 2).

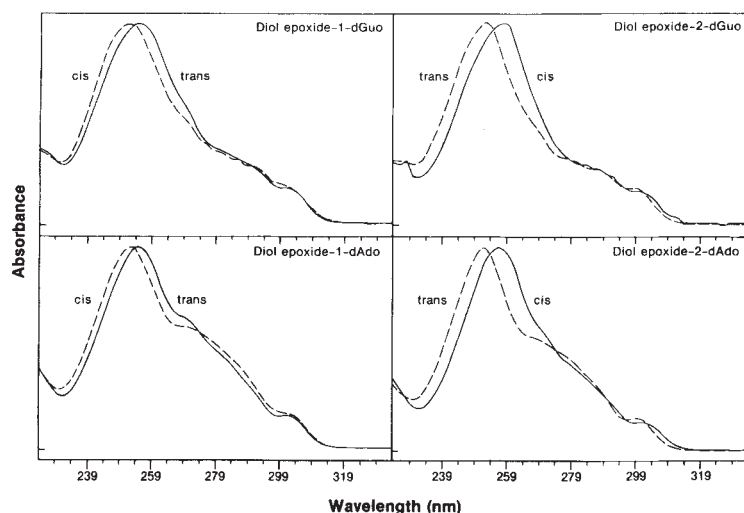


Fig. 2 UV absorption spectra of deoxyribonucleoside adducts derived from DNA and deoxyribonucleotide reactions with the four configurationally isomeric BcPh 3,4-diol-1,2-epoxides. The spectra were obtained with a diode array detector and were measured in the solvents indicated in Fig. 1: 52% methanol for the diol epoxide-1 adducts and 47% and 48% methanol for the diol epoxide-2 adducts. Solid lines, spectra of each dGuo₁ or dAdo₁ adduct; broken lines, spectra of each dGuo₂ or dAdo₂ adduct.

Larger quantities of the 16 purine nucleoside adducts were prepared by reaction of racemic diol epoxides-1 and -2 with purine nucleotides. Detailed characterization of these nucleotide-derived adducts has established that they result from *cis* and *trans* opening of the oxirane ring at C₁ by the exocyclic amino groups of adenine or guanine¹⁴. For diol epoxide-1, the adducts resulting from *trans* opening are dGuo₁ and dAdo₁, and those from *cis* opening are dGuo₂ and dAdo₂. However, in the case of diol epoxide-2, these assignments are reversed because *cis* adducts (dGuo₁ and dAdo₁) elute before the *trans* adducts (dGuo₂ and dAdo₂) (Fig. 2).

This first direct comparison of DNA reaction products from the four configurational isomers of a bay-region diol epoxide indicates that each isomer produces a unique product distribution from DNA. For example, (+)-diol epoxide-1 exhibits a great preference for reaction with deoxyadenosine residues in DNA (the ratio of dAdo to dGuo adducts is ~8.5) and is opened preferentially *trans* by this nucleoside (*trans*:*cis* ratio ≈ 3). The other three diol epoxides are distributed more evenly over both deoxyguanosine and deoxyadenosine residues (dAdo:dGuo ratio ranges between 1.4 and 2) but (–)-diol epoxide-1 is preferentially opened *trans* by deoxyguanosine and *cis* by deoxyadenosine residues. Both diol epoxide-2 enantiomers preferentially form *trans* products, though this preference is most exaggerated in the case of the (–)-enantiomer. Approximately the same extent of adduct formation occurred with deoxyguanylic and deoxyadenylic acids. Both diol epoxide-1 enantiomers showed a preference for *cis* opening of the oxirane ring (*cis*:*trans* ratio ≈ 3.8 and 3.5 for (+)-1 and (–)-1 reactions with deoxyguanylic acid and ~1.7 and 5.3 for (+)-1 and (–)-1 reactions with deoxyadenylic acid) whereas the diol epoxide-2 enantiomers exhibited an even greater preference for *trans* opening (*trans*:*cis* ratio ≈ 16.9 and 9.4 for (+)-2 and 6.3 and 12.2 for (–)-2 with deoxyguanylic and deoxyadenylic acids respectively). DNA, therefore, exerts a particularly dramatic influence on the reactions of (+) diol epoxide-1 in that, in contrast to the nucleotides, it strongly favours reaction with adenine over reaction with guanine and it causes preferential *trans* opening.

The remarkable features of these diol epoxide-DNA reactions are the efficiency of covalent binding to DNA relative to DNA-catalysed hydrolysis and the extensive reaction of these diol epoxides with deoxyadenosine residues. Extensive reaction with deoxyadenosine residues has been noted before only for DMBA^{6–8} which, like BcPh, possesses a sterically crowded bay region. This crowding is accommodated by large out-of-plane distortions^{15,16} which may be associated with high reactivity of the diol epoxide derivatives towards deoxyadenosine residues in DNA. Although our product distributions are derived from

studies *in vitro* rather than *in vivo*, the only adduct which varies in a similar fashion to the tumour-initiating activities¹¹ of these diol epoxides is the deoxyadenosine adduct in which the epoxide ring is opened *trans* (dAdo₁ for diol epoxide-1 and dAdo₂ for diol epoxide-2). Thus, this is a minor product for the least active isomer, (–)-diol epoxide-1, is present in moderate yield for the more active (+)-diol epoxide-2 and is the major product for the two most active isomers, (–)-diol epoxide-2 and (+)-diol epoxide-1. These latter are the most active diol epoxide tumour initiators for mouse skin known¹¹. This relationship for the diol epoxides, together with the extensive reaction with adenine residues for the most potent tumour initiator amongst the parent hydrocarbons, DMBA, suggests a strong association between reactivity towards adenine in DNA and tumour-initiating activity. In DMBA induced tumours, *ras* oncogenes have been found to have been activated by mutational changes in adenine residues at codon 61 (refs 17–19). For BaP diol epoxide activation of the *ras* proto-oncogene *in vitro*, adenine in codon 61 is again the most efficient target for activation because 36% of the activating mutations are at adenine whereas <6% of the total reaction is with this base²⁰. Thus, for hydrocarbon carcinogens, this evidence supports the concept that an interaction with DNA can lead directly (rather than indirectly) to the activation of the *ras* oncogene and to tumour initiation.

This research was sponsored in part by the NCI, DHHS, under a contract with Bionetics Research, Inc.

Received 3 February; accepted 10 April 1987.

1. Sims, P., Grover, P. L., Swaisland, A., Pal, K. & Hewer, A. *Nature* **252**, 326–328 (1974).
2. Buening, M. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 5358–5361 (1978).
3. Meehan, T., Straub, K. & Calvin, M. *Nature* **269**, 725–727 (1977).
4. Ashurst, S. W. & Cohen, G. M. *Chem. Biol. Interact.* **29**, 117–127 (1980).
5. Lo, K.-y. & Kakunaga, T. *Biochem. biophys. Res. Commun.* **99**, 820–829 (1981).
6. Dipple, A., Pigott, M. A., Moschel, R. C. & Costantino, A. *Cancer Res.* **43**, 4132–4135 (1983).
7. Dipple, A., Sawicki, J. T., Moschel, R. C. & Bigger, C. A. H. in *Extrahepatic Drug Metabolism and Chemical Carcinogenesis* (eds Rydstrom, J., Montelius, J. & Bengtsson, M.) 439–448 (Elsevier, Amsterdam, 1983).
8. Bigger, C. A. H., Sawicki, J. T., Blake, D. M., Raymond, L. G. & Dipple, A. *Cancer Res.* **43**, 5647–5651 (1983).
9. Jerina, D. M. & Daly, J. W. in *Drug Metabolism* (eds Parke, D. V. & Smith, R. L.) 13–32 (Taylor & Francis, London, 1976).
10. Levin, W. *et al. Cancer Res.* **40**, 3910–3914 (1980).
11. Levin, W. *et al. Cancer Res.* **46**, 2257–2261 (1986).
12. Jerina, D. M. *et al. in Biological Reactive Intermediates III* (eds Snyder, R., Kocsis, J. J., Jollow, D. J. & Witmer, C. M.) 11–30 (Plenum, New York, 1986).
13. Moussaoui, K., Geacintov, N. E. & Harvey, R. G. *Biophys. Chem.* **22**, 285–297 (1985).
14. Agarwal, S. K. *et al. J. Am. chem. Soc.* **109**, 2497–2504 (1987).
15. Zacharias, D. E. *et al. Carcinogenesis* **5**, 1421–1430 (1984).
16. Hirshfeld, F. L. *J. chem. Soc.* 2126–2135 (1963).
17. Zarbl, H., Sukumar, S., Arthur, A. V., Martin-Zanca, D. & Barbacid, M. *Nature* **315**, 382–385 (1985).
18. Quintanilla, M., Brown, K., Ramsden, M. & Balmain, A. *Nature* **322**, 78–80 (1986).
19. Dandekar, S., Sukumar, S., Zarbl, H., Young, L. J. T. & Cardiff, R. D. *Molec. cell. Biol.* **6**, 4104–4108 (1986).
20. Vousden, K. H., Bos, J. L., Marshall, C. J. & Phillips, D. H. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1222–1226 (1986).

Growing cultures in perfusion bioreactors

Y. Fouron

Perfusion bioreactors offer a culture environment similar to in vivo conditions for mammalian cells, and have already gained commercial importance.

THE recent literature is filled with announcements of newly cloned and expressed proteins. In the past several years, medical publications have reported a stream of successful clinical applications for monoclonal antibodies and recombinant proteins, many of which require mammalian cell expression systems. The rapidly growing need for large amounts of these proteins, for pharmaceutical as well as industrial uses in cell culture media and large-scale immunopurification, has made clear the need for efficient production technologies, and has pointed up the severe limitations of batch fermentation systems as a method for animal cell culture.

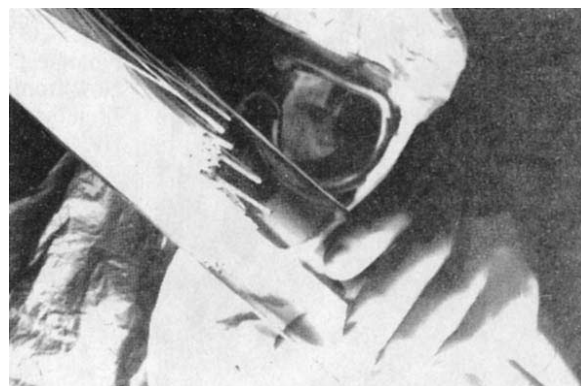
Continuous perfusion bioreactors have been developed to accommodate the special needs and unique characteristics of animal cells. They mimic *in vitro* the conditions under which mammalian cells live *in vivo*, and produce large amounts of proteins at acceptable production costs: an unachievable goal using bacterial fermentation technology.

Fastidious characters

Animal cells have several characteristics that must be taken into account in planning a production technology. First and foremost, they are fragile, and the shear forces present in a bacterial fermenter can destroy them. They survive and produce at optimum levels when immobilized and in close proximity, at cell densities on the order of 10^{10} cells per cubic centimeter. Further, the productivity of animal cells in culture is very low, whether the cell secretes the protein naturally or is genetically engineered, and often reaches only a few micrograms for 10^6 cells grown over 24 hours. Therefore, large numbers of cells, maintained for long periods of time, are needed to manufacture a significant amount of product.

The average animal cell doubling time, from 20 to 40 hours, is also slow compared to bacteria, which double every 20 to 30 minutes. It takes a long time between the inoculation of a bioreactor at 10^5 cells cm^{-3} and the point at which production density, $10^7 - 10^8$ cells cm^{-3} , is reached. Another limiting factor is the existence of negative feedback mechanisms that shut off production or secretion if the concentration of the product is too high in the surrounding media. Finally, the nutritional requirements of animal cells are not well understood. Typical media are complex and require added proteins, often in the form of fetal calf serum, of undefined activities.

Fig. 1 The individual fibres comprising a typical perfusion bioreactor are shown being fitted into a cartridge.



Continuous perfusion is the only approach currently available that can solve the production limitations imposed by these characteristics. It allows for cells to grow to densities of 10^8 to 10^9 cells per cubic centimeter for periods of time extending up to a year. With hybridomas, this results in reactor productivities of tens of milligrams per hour per litre of reactor volume, or hundreds of grams of monoclonal antibodies per year for a 1 litre bioreactor. The fragile protein products can also be removed continuously, limiting their exposure to the harsh elements of the bioreactor environment such as relatively high temperatures and the presence of proteases, thereby increasing the yield of biologically active material.

Perfusion bioreactors

Hollow fibre bioreactors, consisting of bundles of hollow fibres contained within a cylinder, were pioneered by Knazek^{1,2} in the early seventies for use with both suspension and attachment-dependent cells³. The fibres are open at both ends, creating two compartments in the unit: the interior space of the fibres, or luminal space, and the extracapillary space on the outside of the fibres. The permeability of the polymers composing the hollow fibres leads to selective exchange and retention of molecules between fibres and the extracapillary space. The fibre bundle is kept in place by potting at both ends with silicone-rubber materials or various resins.

Cells grow in the extracapillary space while medium is pumped through the cartridge. Media escapes through the fibre wall into the extracapillary space while travelling down the bore, bringing dissolved gases and nutrients to the cells. In the reverse flow, low molecular weight metabolites move toward the inside of the fibres and exit the unit through the lumen. Large product molecules, such as antibodies with molecular weights in the

150,000 to 1,000,000 range, cannot escape from the extracapillary compartment.

The protein products can be collected by draining the extracapillary space, or a continuous flow mechanism can be set up, allowing the product to be processed online. Either way, the product is recovered at a high concentration, decreasing the total cost of downstream processing needed to increase purity.

Several problems must be overcome in using a perfusion bioreactor, however. As a result of the differences in hydrostatic pressure, an axial and radial gradient is established in the unit. To counter this, an optimal ratio of fibre length to diameter should be determined. The flow of media also tends to push the cells toward the exit end of the cylinder, and this can be corrected by reversing the flow periodically.

The ratio of luminal surface area to the extracapillary volume is of prime concern. Over 30 years ago, it was noted in neoplastic tissues that necrosis from lack of oxygen resulted if the distance between the farthest cells and the nearest capillary was greater than 150 μm . As a general rule, this corresponds to a 10-cell-deep layer. If capillaries with an outside diameter of 300 μm are tightly packed in a cylinder, the majority of cells in the extracapillary compartment will be within the required 150 μm distance from a fibre. Most commercial fibres used in cell culture systems have inside diameters ranging from 200 to 300 μm , with wall thicknesses of between 50 and 70 μm , for a total surface area of several square meters.

A look to the future

High productivity plateaus, reflecting the achievement of a steady state, can now be maintained in perfusion bioreactors. Higher cell densities and higher secretion per cell may be reached in future through the further development of high performi-

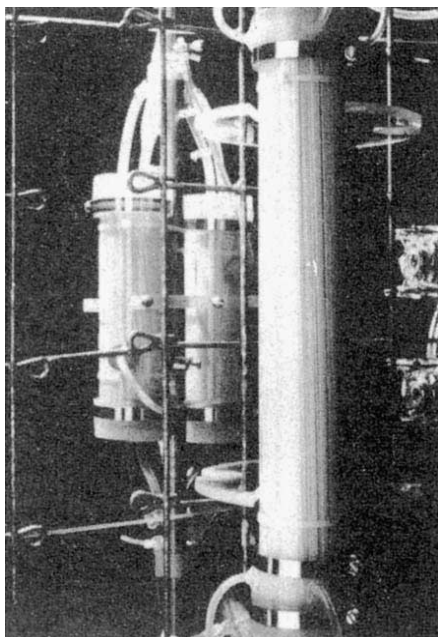


Fig. 2 A working perfusion bioreactor setup.

ance chemically defined media containing cloned growth factors.

Improvements in online monitoring will be attained by the development of sensitive biosensors that can measure accurately and selectively molecules secreted or removed from media by cells in culture. The first generation of these biosensors are just arriving on the biotech scene⁴, although very few are now in use. The application of artificial intelligence has been proposed for the design and operation of bioreactors⁵, once accurate monitoring becomes possible. The coming decade will see the integration of these advances as well as the development of "smart membranes" based on new polymers and fabrication techniques⁶. In the last few years the relationship between monomer structure and the physico-chemical properties of the resulting polymer have become more clear. Composite polymers, high performance blends, and post-polymerization surface treatments promise to deliver custom-designed membranes with selective permeability, and membranes with a permeability that can be electrically modulated. The first case of synthetic polymer self-assembly⁷ takes science one step closer to the goal of making polymeric materials with a high resemblance to natural biological membranes. □

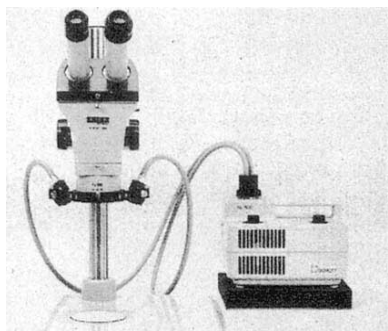
Yves Fouron is at Bio-Response, 1978 West Winton Avenue, Hayward, California 94545, USA. For more information, fill in reader service number 100.

1. Knazek, R.A., Gullino, P.M., Kohler, P.O., & Dedrick, R.L. *Science* **178**, 65-68 (1972).
2. Gullino, P.M. & Knazek, R.A. *Meth. Enzym.* **58**, 178-184 (1979).
3. Tharakan, J.P. & Chau, P.C. *Biotech. Bioengng* **18**, 329-342 (1986).
4. North, J.R. *Trends Biotech.* **3**, 180-186 (1985).
5. Stephanopoulos, G. & Stephanopoulos, G. *Trends Biotech.* **4**, 241-249 (1986).
6. Baer, E. *Scient. Am.* **255**, 178-190 (1986).
7. Yoon, H.S. *Nature* **326**, 580-582 (1987).

Biotechnology on display

Next week, exhibitors from all over Europe will converge on Amsterdam, the Netherlands, for the 4th European Congress on Biotechnology, ready to show off their bioreactors, chromatographs, and products for molecular biology.

ZEISS will be displaying its line of research biological microscopes in booth 132. New from Zeiss is a Dual Incident-Light Illuminator system for stereomicroscopes (Reader Service No. 101). The light guides of the \$1,536 (US) **stereomicroscope illuminator** attach directly to the objective dovetail, so the illumination is always adjusted to the focal plane, even when the height of the specimens under



The Zeiss illuminator system sheds new light on stereomicroscopy.

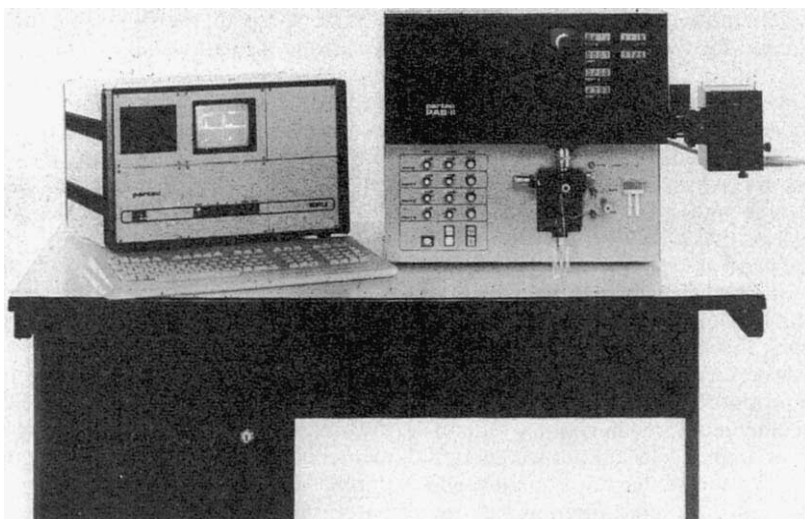
examination vary. Zeiss says the fibre light cables are flexible, and offer virtually no resistance when adjusting the microscope. Each of the illuminators can be focused individually, and brightness, illumination angle and size of the luminous field can be controlled separately. Zeiss recommends the Schott KL 1500 light source, with variable brightness between 6.3 and 100 per cent in five steps, for use with the illuminator system.

The PAS II flow cytometer and cell sorter from Partec can analyse the parameters important for cell growth,

such as DNA, RNA or protein content, of up to 1,000 cells per second, according to the company (Reader Service No. 102). The Partec system uses a closed, fully sterilizable system, making it possible to sort cells, chromosomes or microorganisms for further experimentation. The system's computer can be programmed to sort cells according to independently selectable criteria. The basic PAS II system carries a SF40,000 (Switzerland) price tag, or the system can be purchased with additional options for SF170,000 (Switzerland). More information will be available from Partec in booth 144.

Biotech Instruments Ltd, exhibiting in booth 143/8, is distributing Hoefer Scientific's ProGenetor **electroporation systems** (Reader Service No. 103). The ProGenetor systems deliver pulses between 10 μ s to 999 ms long, with nearly 200 incremental settings. Because voltage integral has been shown to be more effective than a high voltage level for electroporation, the Progenetor gives relatively long pulses at low voltages of between 25 and 500 volts d.c. The single ring electrode produces a voltage gradient of 40-800 V cm^{-1} , and the triple ring electrode gives a range of 180-3,500 V cm^{-1} . Both electrodes can be flame-sterilized, and fit into standard 24-well tissue culture trays. Biotech Instruments prices the ProGenetor systems at £1,100-1,400 (UK).

In booth 147, Bio-Rad will be displaying its range of **products for biotechnology** and molecular biology. New developments include the Sequi-Gen nucleic acid

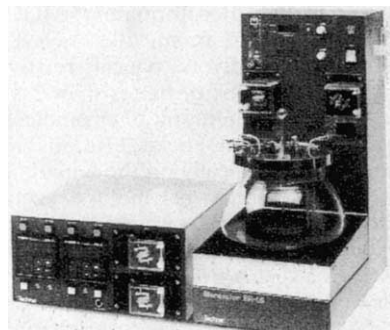


Partec's PAS II flow cytometer quantifies and sorts a wide range of microorganisms.

sequencing system, the Gene Master sequence reader and Gene Pulse transfection apparatus. To supplement these offerings, Bio-Rad also has a line of electrophoresis and ELISA equipment, HPLC systems and column chemistries for monoclonal antibody production. Bio-Rad's free catalogue (Reader Service No. 104) describes these products, with five sections covering DNA sequencing, molecular cloning, nucleic acid blotting, electrophoresis and chromatography.

Bioreactors that produce

Techne, in booth 143/9 has enlarged and improved its **floating impeller bioreactor**, to produce the BR-06 Tecam (Reader Service No. 105). The Tecam bioreactor vessel handles working volumes of between 500 and 3,000 ml, and is composed of borosilicate glass that is siliconized to keep cells from attaching to and growing on the reactor wall. The spherical vessel has a wide mouth and a fitted stainless-



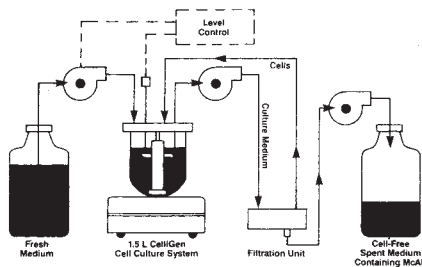
The floating impeller of Techne's Tecam bioreactor aids oxygen exchange.

steel lid with ten ports for media addition, gassing, sampling pH and pO_2 , and other uses. The Techne Tecam is suitable for mammalian and plant suspension cultures, and can also accommodate anchorage-dependent cell lines grown on microcarriers. The floating magnetic impeller agitates the surface to facilitate oxygen transfer, eliminating the bearings and electric drive motor usually attached to fermenter lids. Techne says hybridoma growth profiles show that the \$5,800–7,300 (US) Tecam bioreactor maintains a higher concentration of viable cells than either a T-75 flask or an airlift fermenter.

The focus of LH Fermentation's 2000 Series of **fermenter systems** is versatility (Reader Service No. 106). The 2000 Series mainframe is designed to accommodate a wide range of instrumentation and a range of interchangeable vessels. A modular analogue controller capable of computer setpoint control regulates the series, in either batch or continuous culture mode. The 2000 Series I fits glass reactor vessels of 3.5, 5, 7, 8, 12, 16 and 20 litre capacities, that are designed for *in situ* or autoclave sterilization. The 2000 Series II is scaled up for pilot plant research needs, and features stainless steel vessels of 20, 42,

and 70 litre volumes in addition to the 3.5 to 20 litre glass vessels. LH Fermentation will be showing its fermentation products in booth 143/4.

New Brunswick Scientific, exhibiting in booth 140, has just unveiled its new **Monoclonal Antibody Factory benchtop**



A schematic of New Brunswick's MAb Factory shows the setup for maximum MAb production.

perfusion system, that is designed to produce over 500 mg l^{-1} per day (Reader Service No. 107). The centerpiece of the MAb Factory is New Brunswick's CelliGen cell culture bioreactor, that it reports can achieve suspension culture cell densities in excess of 1.8×10^7 cells ml^{-1} with 90 per cent viability. According to the company, in the growth of mouse-mouse hybridoma, antibody concentrations have reached 800 $\mu g\ ml^{-1}$ in batch culture and 504 $mg\ l^{-1}$ in continuous production, yielding 7.8 g of product over 48 days. The CelliGen bioreactor has a low-shear gas exchange impeller and a cell lift impeller that generates a suction-lift when rotated gently. Culture fluid is passed through a hollow fibre cartridge that draws off the antibody as a cell-free filtrate, and when coupled with New Brunswick's affinity supports, active monoclonal antibodies can be obtained. The entire setup can be purchased as a complete system, or be bought as individual components for existing systems.

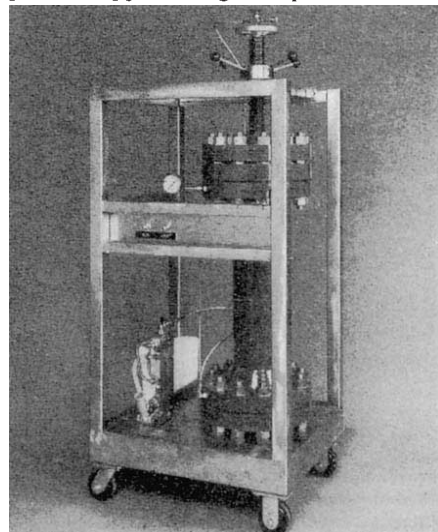
Whatman's biotechnology products are geared toward increasing efficiency in **downstream bioprocessing** (Reader Service No. 108). The Whatman exhibit, in booth 143/13 is likely to contain information on ion-exchange media, process columns and mammalian bioreactors. Whatman says its ion-exchange cellulose media grades offer faster kinetics and higher protein capacity in many processes than biopolymer separations, and its process columns are designed for simple pump packing techniques and high throughputs. Microprocessor-controlled mammalian cell bioreactors are the latest addition to Whatman's line.

For supervising and controlling large-scale fermentation processes, Biotechnology Computer Systems, in booth 143/4, has developed its **Bio-i bioprocessing software** package (Reader Service No. 109). Bio-i runs on DEC computers, and

provides for process management and data collection, storage, processing and presentation. It interfaces not only with a wide range of fermenters, but also collects data from measurement equipment, such as mass spectrometers, associated with the process. Built-in privilege protection codes prevent loss or modification of batch data, and lock out unauthorized personnel. Alarm systems alert process supervisors in the event of deviations from programmed conditions, and remote terminal access options are available for home supervision of production. The software provides plant and batch overview, and plots trend analyses in real time.

Chromatography components

A free-standing preparative and **process-scale HPLC module**, the Kiloprep Compression Module 250, is available from Millipore, displaying in booth 146 (Reader Service No. 110). The Millipore module is designed for speed and maximum product recovery. Using the KPCM 250, Millipore says chromatographers can purify large product loads in minutes or hours, rather than days, with yields and purities approaching 100 per cent. The



Designed for high-volume preparative-scale chromatography.

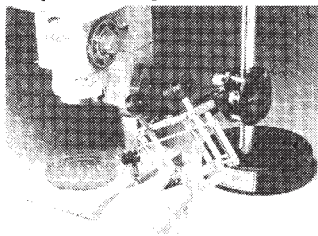
module can be fitted with a four-inch prepacked cartridge that is changed easily, allowing researchers to spend more time doing separations and less time packing and changing columns. The \$25,000 (US) module is compatible with and can be fitted to most of the preparative and process HPLC systems on the market, according to Millipore.

A **diode array detector** has been added to Perkin-Elmer's line of liquid chromatography instruments (Reader Service No. 111). The \$9,950 (US) LC-235 combines diode array with a conventional variable-wavelength UV detector in one unit, to accommodate pharmaceutical, chemical, biomedical and biotechnological needs. Perkin-Elmer says the dual-beam optical

ADVERTISEMENTS

SINGER INSTRUMENTS

Xenopus oocytes injection?



SINGER MKI MICROMANIPULATOR

We also offer a Microsyringe Kit.

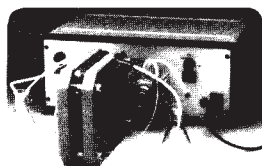
Write to:

Singer Instrument Co. Ltd.
Treborough Lodge
Roadwater, Watchet
Somerset TA23 0QL
England. Tel. (0984) 40226
Telex: 337492 Comcab G
 Circle No.403 on Reader Service Card

Reader Service No.50

LIPOSOMES

REPRODUCIBLE — HOMOGENEOUS —
 EXTREMELY RAPID



Our new LIPOSOMAT guarantees rapid (5ml within 60 minutes) preparation of uniformly sized unilamellar liposomes. Liposome size is selectable between 25 and ca. 600 nm diameter. Sample volume is normally 5-10ml or up to 200ml if used in combination with an additional instrument. Lipid concentration can be up to 300 mg/ml.

DIANORM

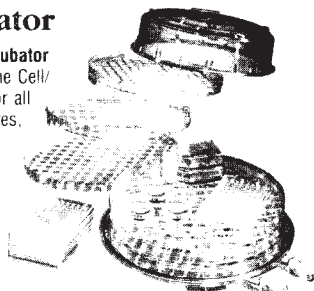
POB 126, D-8000 Munich 65, FRG

Reader Service No.49

Versatile Tissue Culture Incubator

Modular Incubator Chamber (the Cell/Sphere™) for all tissue cultures, aerobic or anaerobic.

Low Cost



Eliminates evaporation loss and contamination. Easy to use — flush with gas mixture, seal, place at appropriate temperature. Stackable, see-through. Quantity discounts. **billups-rothenberg, inc., pob 977, Del Mar, CA USA 92014. Call 619-755-3309 or contact your Flow Laboratories representative.**

Reader Service No.174

Coming in Product Review . . .

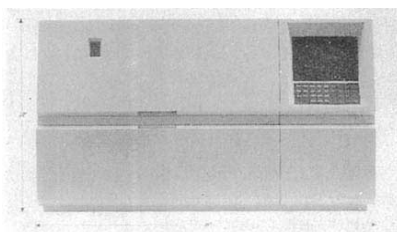
• **Science in Israel.** The 18 June edition of *Nature* will feature a special report on developments and issues in Israeli science, with an accompanying Product Review.

system, covering wavelengths from 195 to 365 nm, and low dispersion flow cell provide low noise, minimum drift and excellent detection limits. Special features include automatic peak purity determination, wavelength ratios and identification of wavelength of maximum absorbance. Perkin-Elmer will be in booth 165.

Racemic drugs can be separated with LKB Produkter's EnantioPac cartridge chromatography columns, allowing therapeutic enantiomers to be isolated, eliminating possible side-effects arising from their inactive partners (Reader Service No. 112). The EnantioPac **chiral column** provides a rapid screening process to test the optical purity of many racemic pharmaceuticals. The chiral stationary phase is comprised of a matrix of silica 10 μ m particles on which the human plasma protein α_1 -acid glycoprotein (α_1 -AGP) has been immobilized. Each molecule of α_1 -AGP provides one chiral interaction site, where temporary diastereomeric complexes form to retard the passage of one enantiomer while the other elutes. The EnantioPac 4 mm i.d. column comes in a reusable holder, and can withstand 100 bar pressure and a maximum temperature of 45°C. LKB Produkter is in booth 118.

Emphasis on proteins

The **protein sequencer** from Porton International, in booth 143/4 was designed to provide the technology for sequencing small amounts of protein to researchers involved in genetic engineering (Reader Service No. 113). The sequencer's small



A bench-top sized machine for small-scale protein sequencing.

size, 36 inches wide $\times$ 16 inches deep $\times$ 22 inches high, lets it fit on the benchtop in most laboratories. It has a menu-driven computer that guides the user through the Edman degradation and ATZ conversion procedures step-by-step, and that also has room for custom programs. The solvent and reagent flasks connect directly to the valve block, to eliminate the potential for tubing leaks or oxygen contamination of sensitive chemicals. Porton's machine costs under £50,000 (UK).

The AminoQuant **amino acid analyser** from Hewlett-Packard automates the analysis of 17 primary and secondary amino acids from the hydrolysates of proteins and peptides (Reader Service



The HP AminoQuant derivatizes amino acids from proteins for further analysis.

No. 114). Both primary and secondary amino acids are derivatized in a pre-column operation in the autosampler/injector system. Primary amino acids are derivatized using a mixture of orthophthalaldehyde (OPA) and 3-mercaptopropionic acid (MPA), while secondary amino acids are derivatized with 9-fluorenylmethylchloroformate (FMOC). HP, exhibiting in booth 102, says the AminoQuant delivers typical relative-standard deviations of better than 2 per cent for measurements on 10 picomoles of all specified amino acids. Results are reported automatically on the amount of amino acid, residues per mole, expected to calculated molecular weight ratios, and percentage of amino acids in the sample.

Chemicals and kits

The IgG RAST **test system** from Pharmacia allows the detection of clinically interesting changes in allergen-specific IgG and IgG₄ in serum during hyposensitization therapy (Reader Service No. 115). The test is based on a sandwich immunoassay that uses monoclonal anti-IgG antibody and a newly developed solid phase to bind the antigen to the test tube wall covalently. The test kit consists of IgG RAST isotope reagents and enzyme-immunoreagents, IgG₄ RAST isotope reagents, reference reagents and allergen tubes. Pharmacia will be in booth 126.

Optical quality **caesium chloride** is being offered by Boehringer Mannheim, exhibiting in booth 123 (Reader Service No. 116). Boehringer Mannheim says the caesium chloride has an absorbance of less than 0.020 (A_{260} , 50 per cent w/v), making it suited for analytical procedures that require absolute minimum UV interference, and costs \$98.50 (US) for 250 g. For less stringent applications such as DNA or RNA purification a lower-priced quality is available at \$35 (US) for 100 g. □

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.